LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Faculty of Business Administration
Munich School of Management

From Open Data to Big Data
Opportunities and Challenges from a Business Perspective

2nd International Open Data Dialog
Berlin
November 18, 2013

Arnold Picot

Research Center for Information, Organization
and Management, Ludwig-Maximilians-University Munich

Münchner Kreis – Non-profit supra-national association
dedicated to communications research

Agenda

1. **Open Data & Big Data – Introduction and Definition**

2. **Open Data & Big Data – Examples of Big Data Applications**

3. **Open Data & Big Data – Selected Challenges**

Agenda

1. **Open Data & Big Data – Introduction and Definition**

2. Open Data & Big Data – Examples of Big Data Applications

3. Open Data & Big Data – Selected Challenges

Digitalization and interconnection lead to an enormous growth of data

- **Digitalization leads to an increasing dematerialization - the transformation of physical objects to electronic information (Picot et al., 2008)**

- **An increased use of media and physical networking through sensor networks for business and private purposes leads to an increasing data generation**

- **This leads to changes in business processes and opens up new business and growth opportunities worldwide**

40 000 000 000

TERABYTE OF DATA

WILL BE GENERATED AROUND
THE WORLD IN 2020
(2012: 2.8 BILLION TERABYTE)
SOURCE: IDC

4,4

MILLION NEW JOBS

IN THE IT SECTOR THROUGH BIG
DATA, WORLDWIDE, BY 2015
SOURCE: GARTNER

4,6

BILLION EURO SALES

WERE GENERATED AROUND
THE WORLD IN 2012 BY
USING BIG DATA
APPLICATIONS.
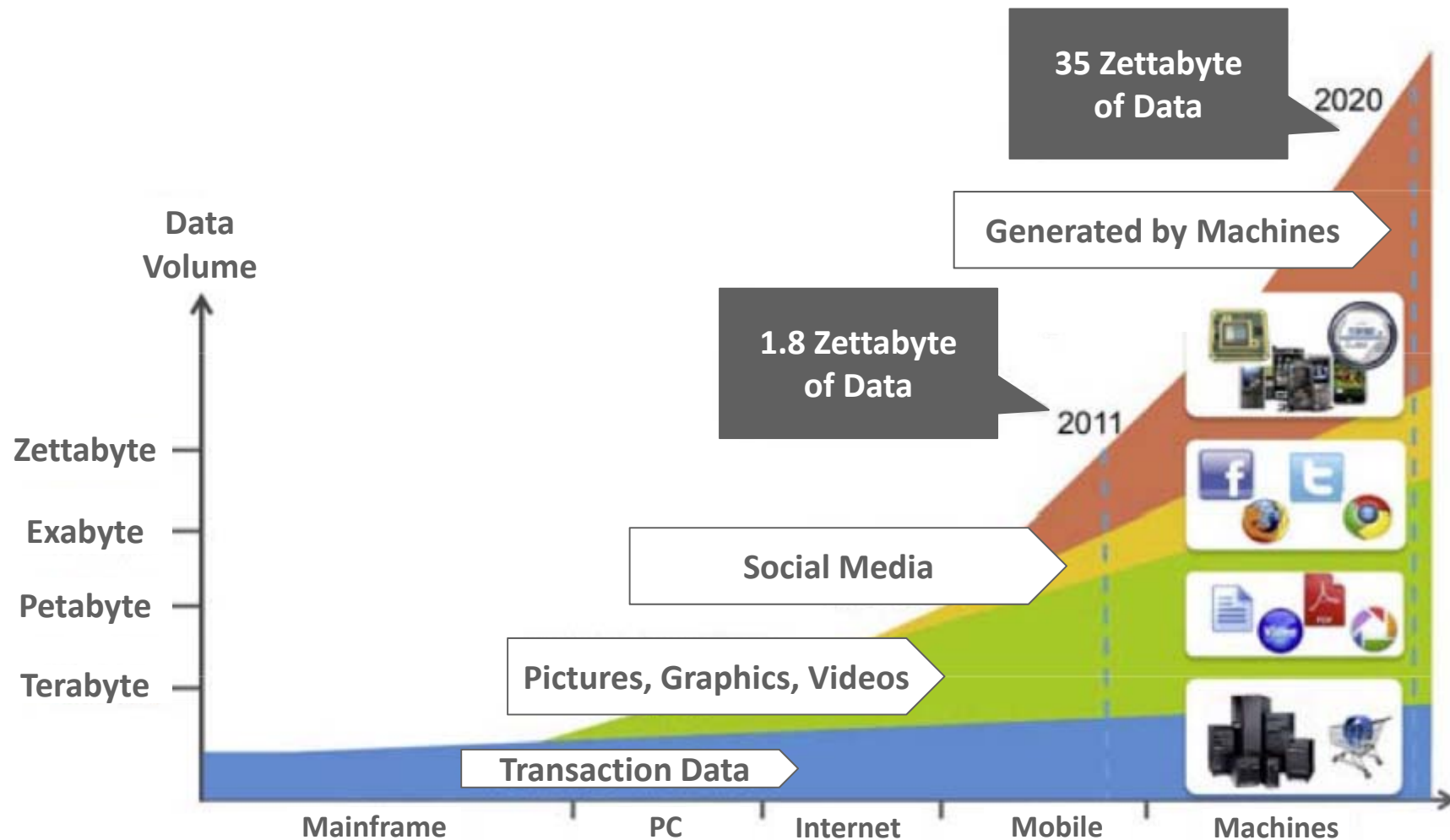By 2016, SALES ARE
SUPPOSED TO REACH
16 BILLIONS IN SALES
SOURCE: BITKOM

Source: Spiegel (2013)

4

The Internet is a key driver for data growth
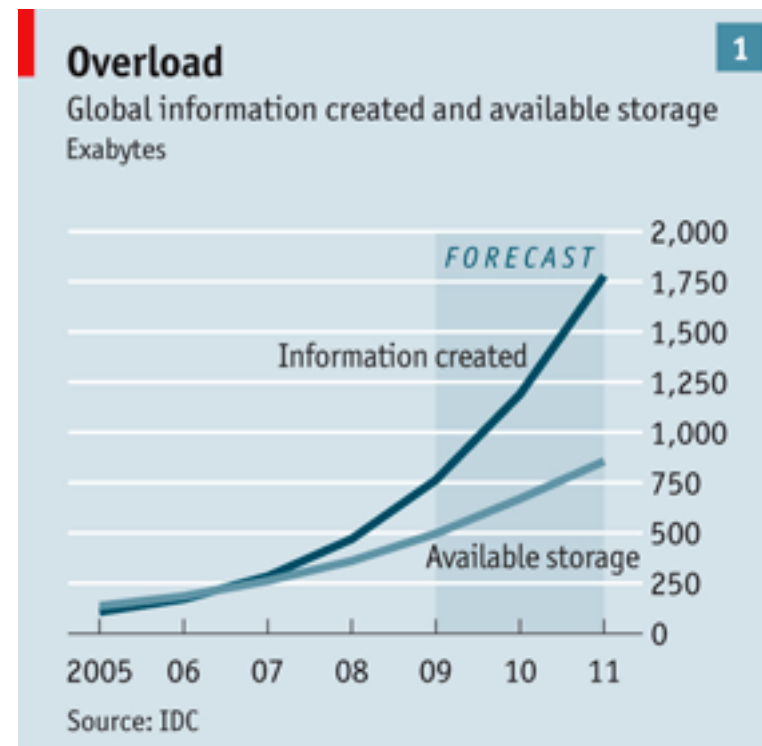
Development of data volume

The flood of data already exceeds the available storage capacity

- **The worldwide generated data already exceed the available storage.**

- **Data sources are particularly:**
  - Motion data (GPS, GSM, etc.)
  - Inventory data
  - Transaction data
  - Sensor data
  - Image, video, audio files
  - Social media data
  - Internet log files

- **What data is worth saving?**

- **Which data can generate value in future?**

Overload
Global information created and available storage
Exabytes

FORECAST
Information created
Available storage

2,000
1,750
1,500
1,250
1,000
750
500
250
0

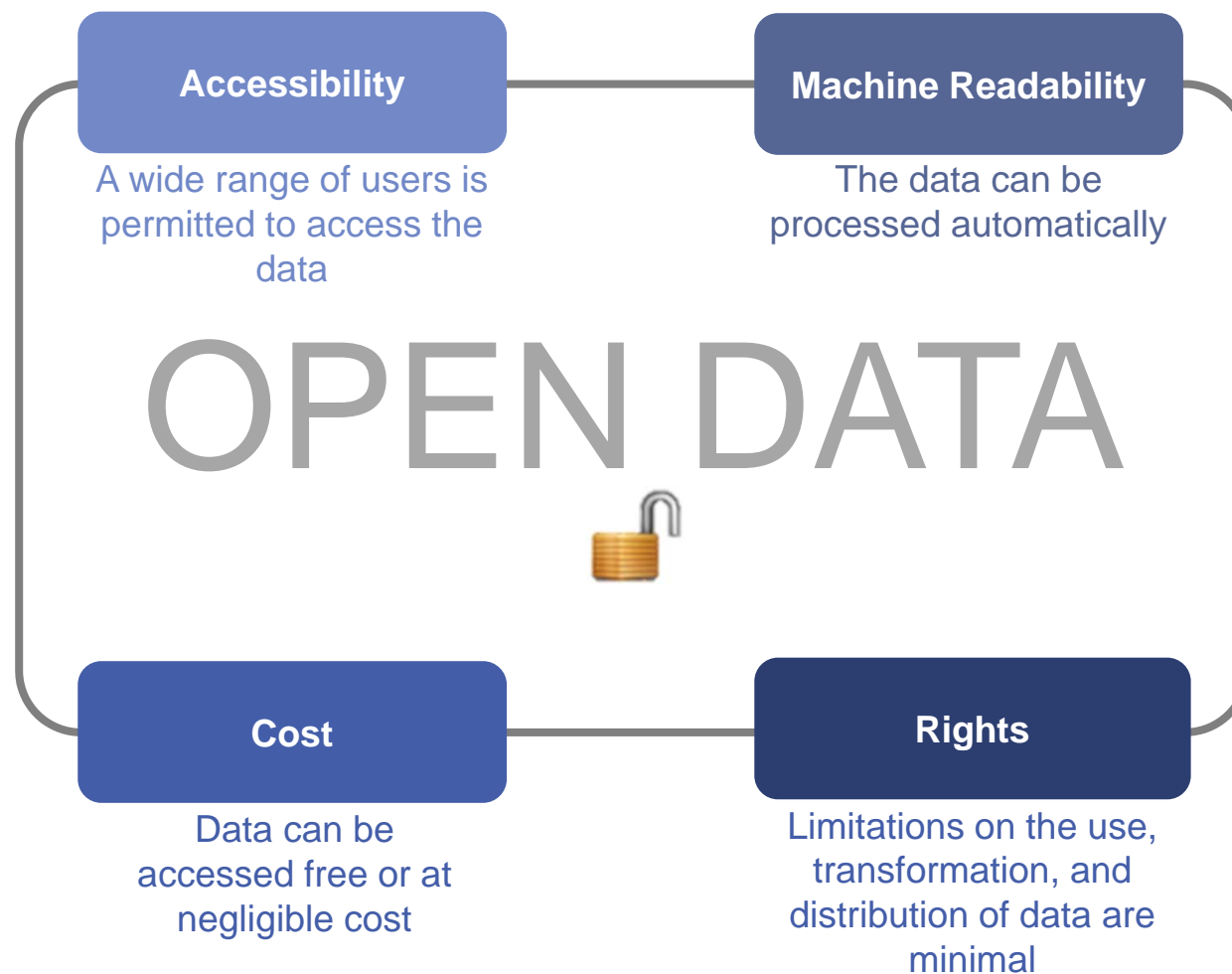2005 06 07 08 09 10 11

Source: IDC

Data as the basis of a digital economy

- **Change of classic business models**

  - Data as key input factor of production lead to huge economies of scale

    - Purpose-related generation, analysis and synthesis of data on product creation is sometimes very expensive (first copy costs)
    - Any further copy or use creates minimal additional costs (second copy costs)

  - Personal data increasingly as (indirect) payment for many services: Facebook, Gmail, XING Skype etc.

"Data is the new oil of the Internet and the new currency of the digital world"
(Meglena Kuneva, European Consumer Commissioner, 2009)

Open Data sets share common characteristics

Accessibility

A wide range of users is permitted to access the data

Machine Readability

The data can be processed automatically

OPEN DATA

Cost

Data can be accessed free or at negligible cost

Rights

Limitations on the use, transformation, and distribution of data are minimal

How data are open or closed, based on four characteristics

Completely Open

More Liquid

Completely Closed

Degree of Access	Everyone has access	Access to data is restricted to a subset of individuals or organizations
Machine Readability	Available in formats that can be easily retrieved and processed by computers	Data in formats not easily retrieved and processed by computers
Cost	No cost to obtain	Offered only at a significant fee
Rights	Unlimited rights to reuse and redistribute data	Re-use, republishing, or distribution of data is forbidden

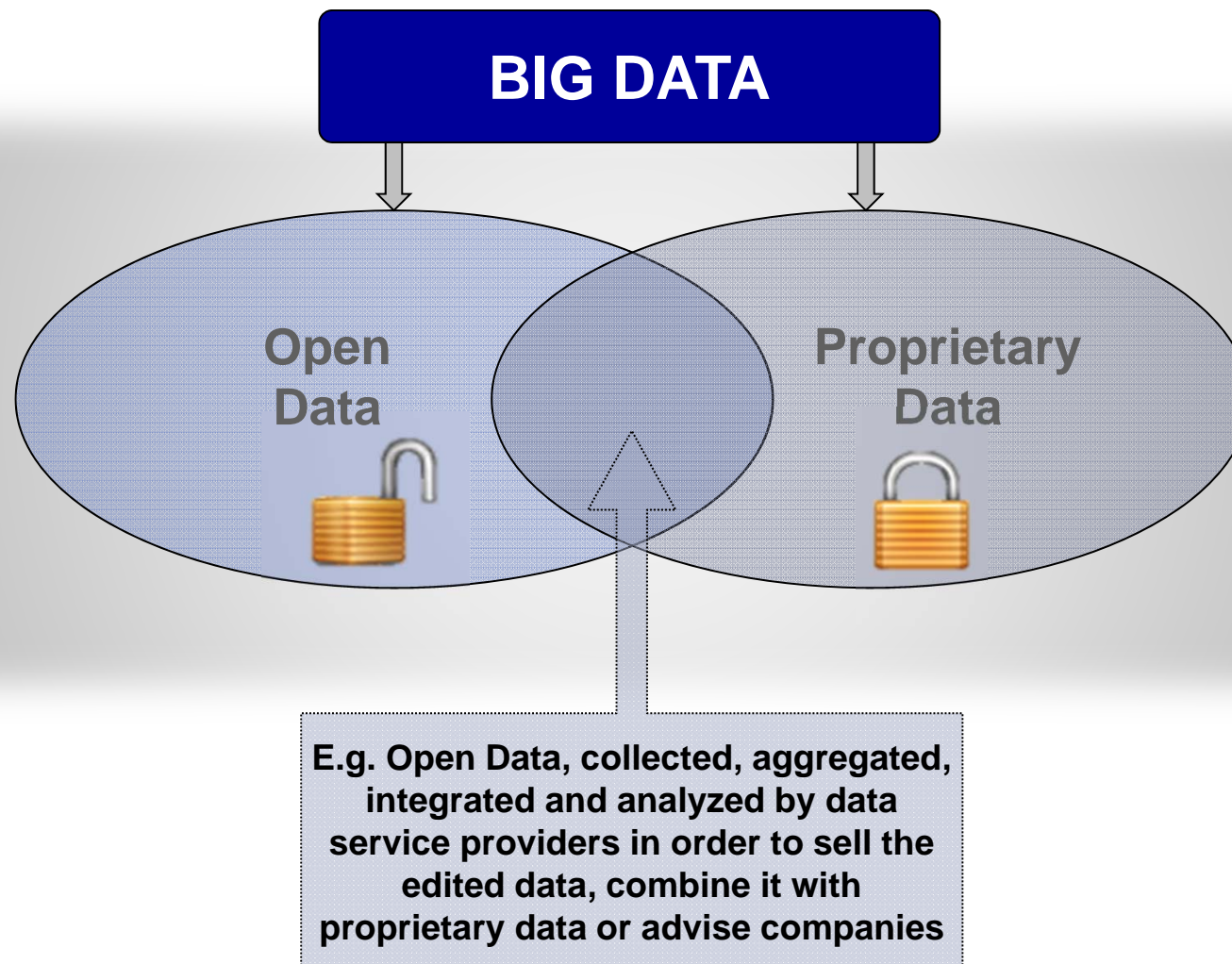Source: McKinsey Global Institute analysis (2013)

Big Data Definition – Big Data is more than just "Big"

3 V's: "Big data is **high-volume**, **-velocity** and **-variety** information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making."
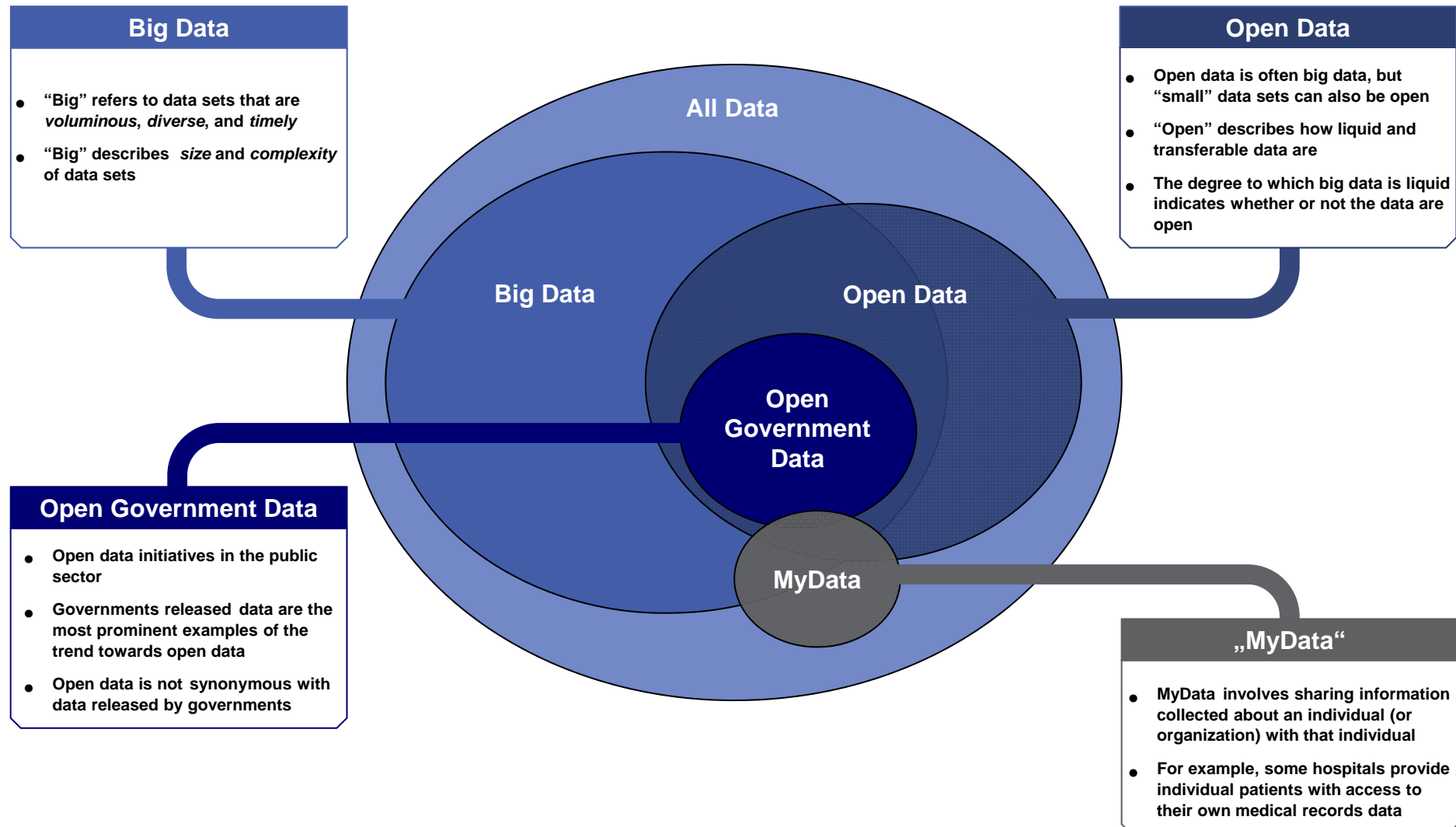(Beyer/Laney, 2012)

GigaByte
TeraByte
PetaByte
ExaByte

Volume — **There are large amounts of data (open and proprietary)**

Text
Audio
Sensor Data
Video
Social Network Data

Variety — **High variety in data sources and formats**

Batch
Periodic
Real Time
Near Real Time

Velocity — **Occurance and analysis of data in (near) real-time**
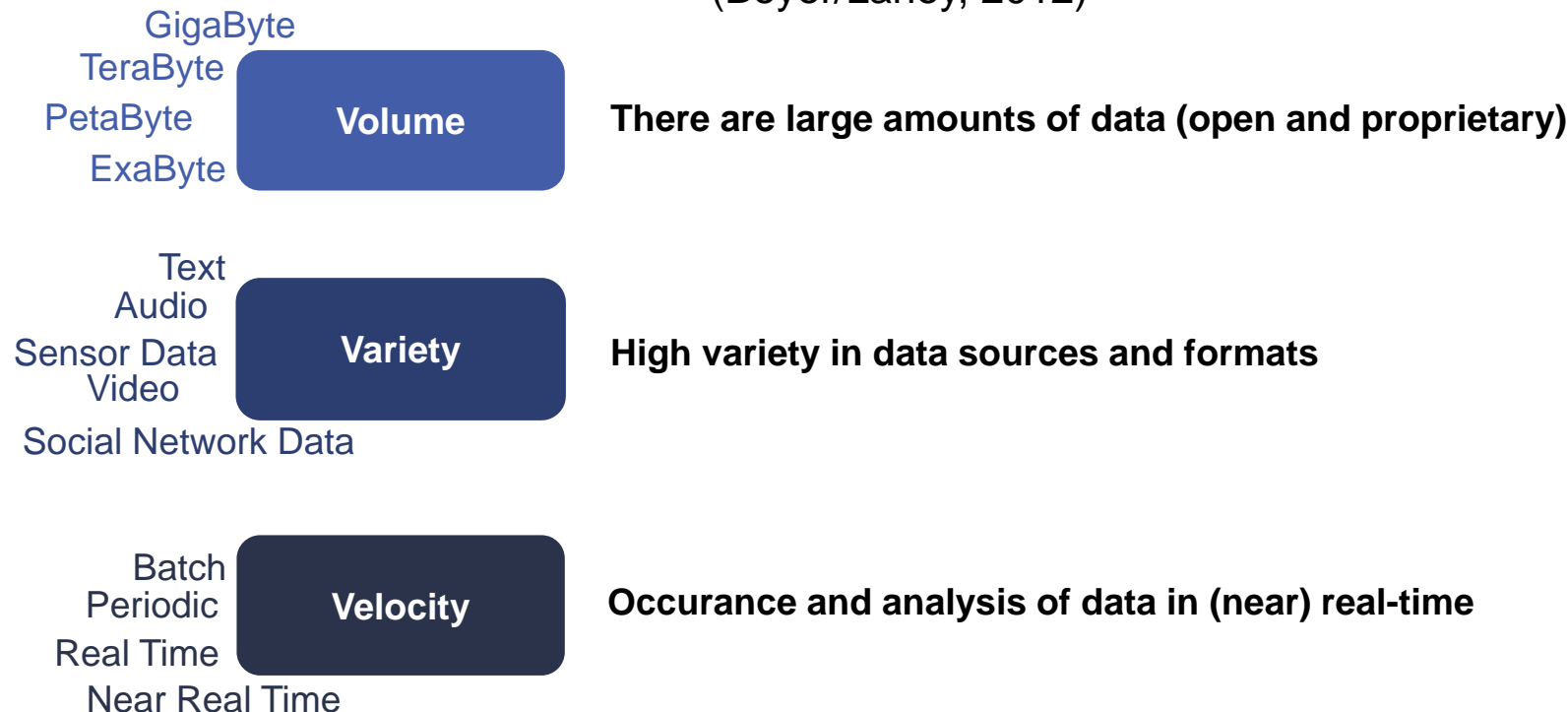
"Big Data is less about data that is big than it is about a capacity to search, aggregate, and cross-reference large data sets." (Boyd/Crawford, 2012)
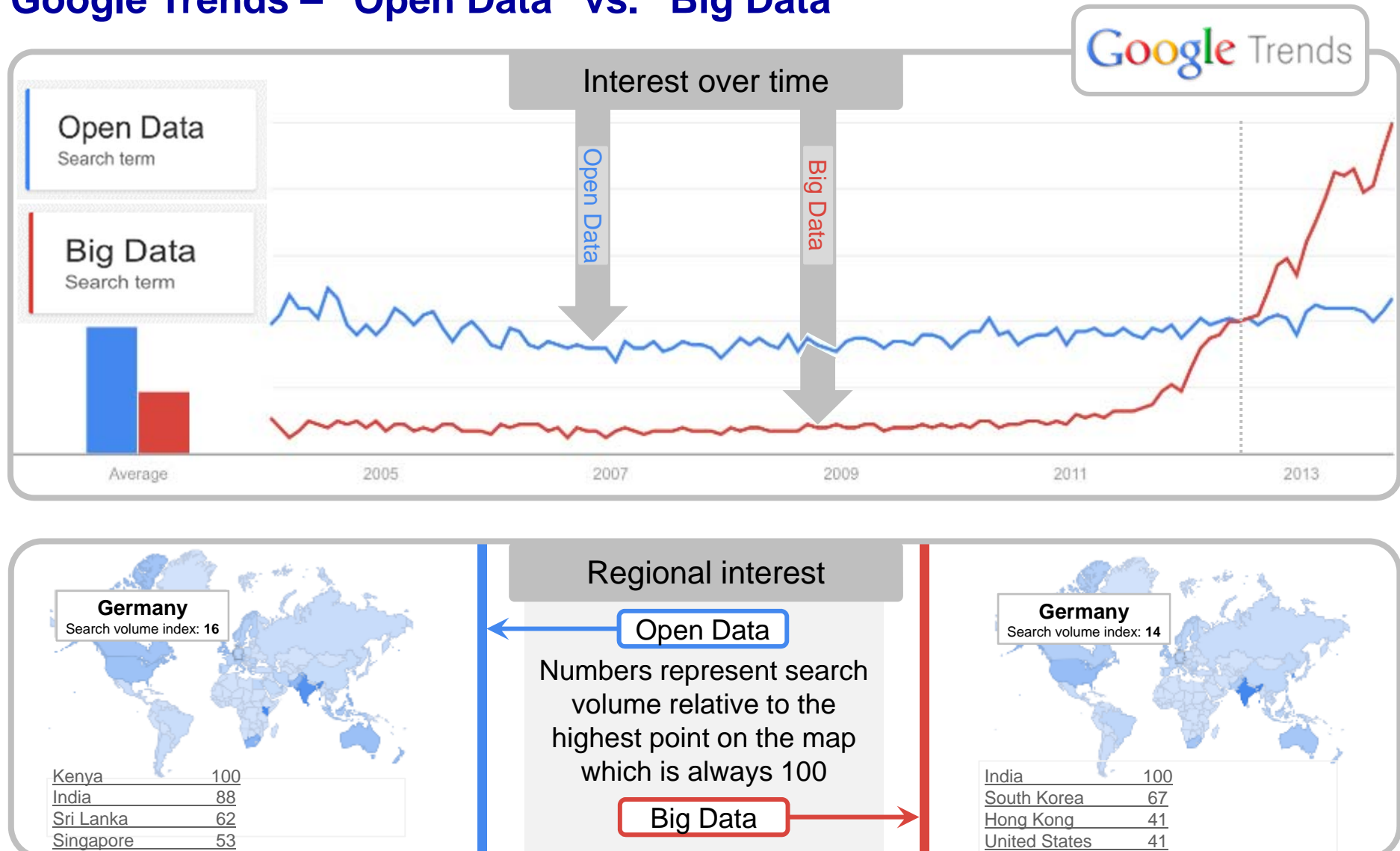
Open Data, Propriety Data, Big Data – A distinction of terms

BIG DATA

Open Data

Proprietary Data

E.g. Open Data, collected, aggregated, integrated and analyzed by data service providers in order to sell the edited data, combine it with proprietary data or advise companies

Source: Own illustration

How Big Data and Open Data relates to other types of data

Big Data

- "Big" refers to data sets that are *voluminous*, *diverse*, and *timely*

- "Big" describes *size* and *complexity* of data sets

Open Data

- Open data is often big data, but "small" data sets can also be open

- "Open" describes how liquid and transferable data are

- The degree to which big data is liquid indicates whether or not the data are open

Open Government Data

- Open data initiatives in the public sector

- Governments released data are the most prominent examples of the trend towards open data

- Open data is not synonymous with data released by governments

„MyData"

- MyData involves sharing information collected about an individual (or organization) with that individual

- For example, some hospitals provide individual patients with access to their own medical records data

All Data

Big Data

Open Data

Open Government Data

MyData

13

Google Trends – "Open Data" vs. "Big Data"

Source: Illustration based on www.google.com/trends/
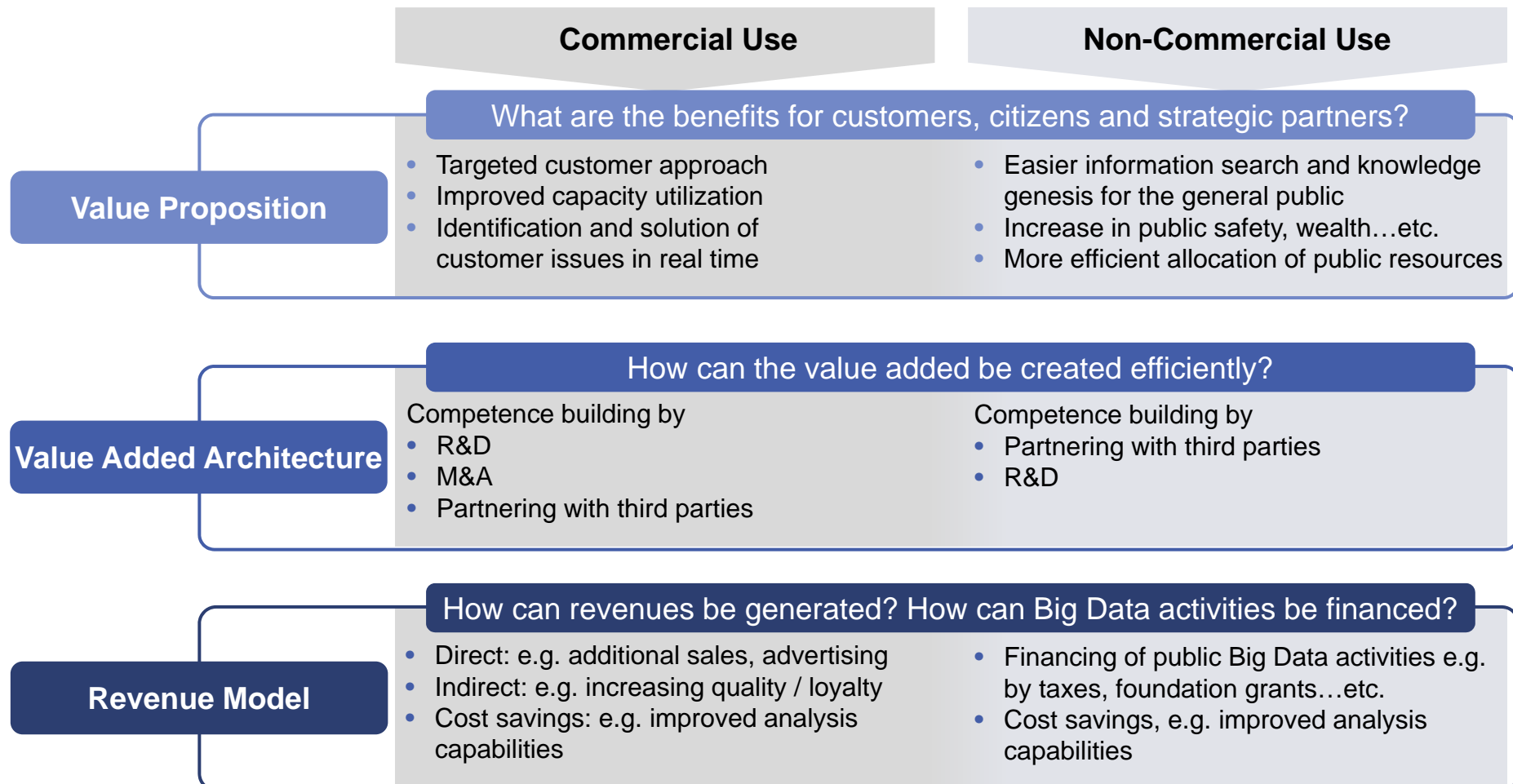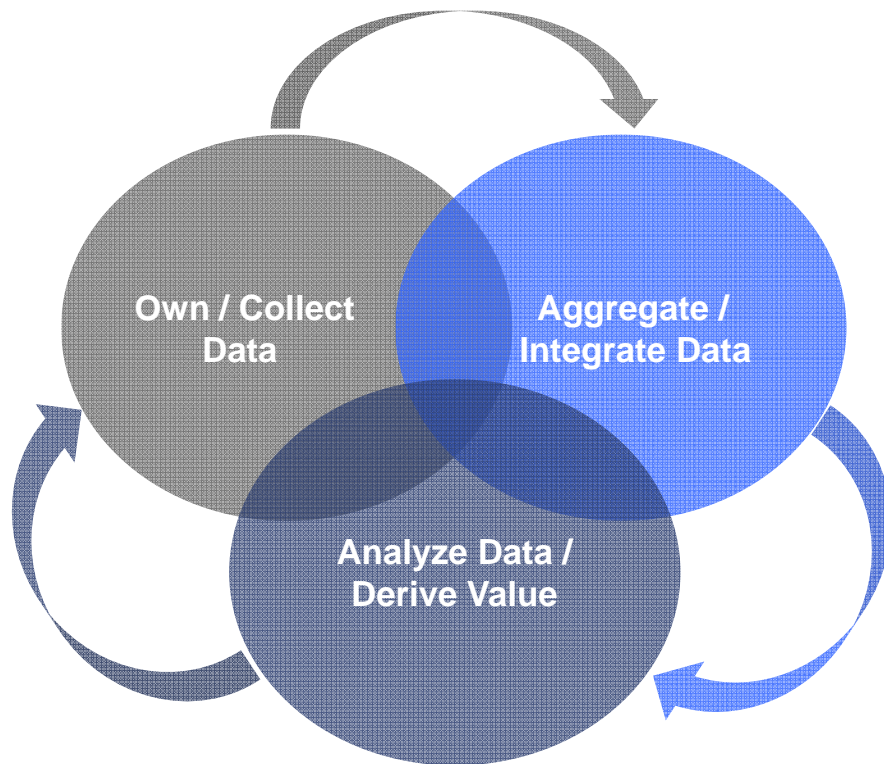
Appropriate business models need to be found

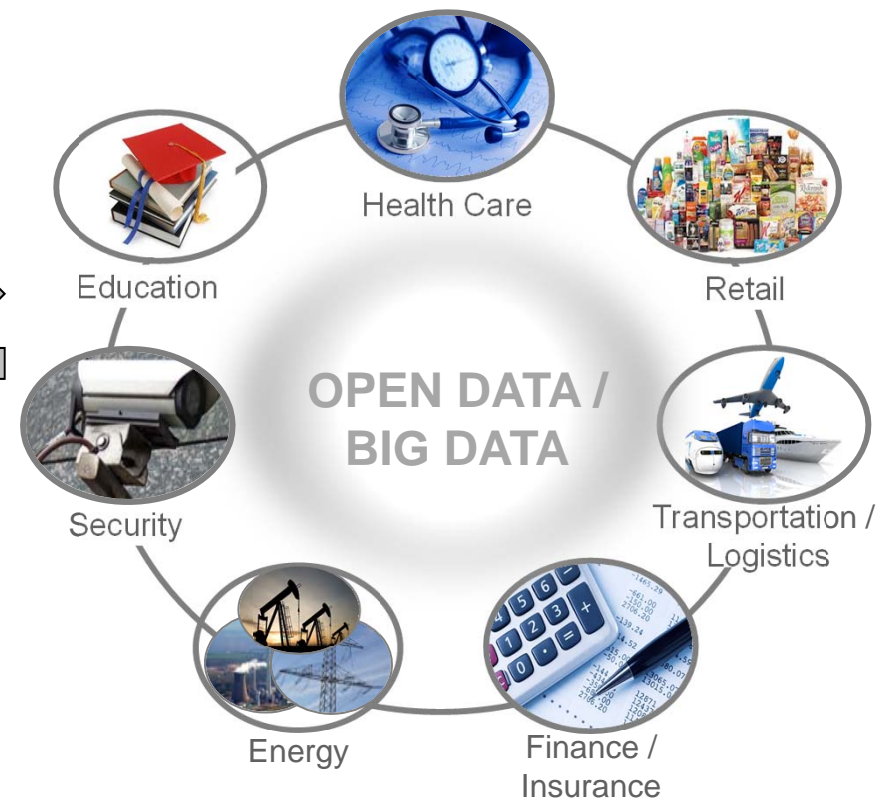## How can Big Data contribute to sustainable advantage for citizens and customers?

	Commercial Use	Non-Commercial Use
Value Proposition	**What are the benefits for customers, citizens and strategic partners?**	
	• Targeted customer approach • Improved capacity utilization • Identification and solution of customer issues in real time	• Easier information search and knowledge genesis for the general public • Increase in public safety, wealth…etc. • More efficient allocation of public resources
Value Added Architecture	**How can the value added be created efficiently?**	
	Competence building by • R&D • M&A • Partnering with third parties	Competence building by • Partnering with third parties • R&D
Revenue Model	**How can revenues be generated? How can Big Data activities be financed?**	
	• Direct: e.g. additional sales, advertising • Indirect: e.g. increasing quality / loyalty • Cost savings: e.g. improved analysis capabilities	• Financing of public Big Data activities e.g. by taxes, foundation grants…etc. • Cost savings, e.g. improved analysis capabilities

Players in the Open Data / Big Data environment: The "Big Data Circus"

Big Data "Value Chain"

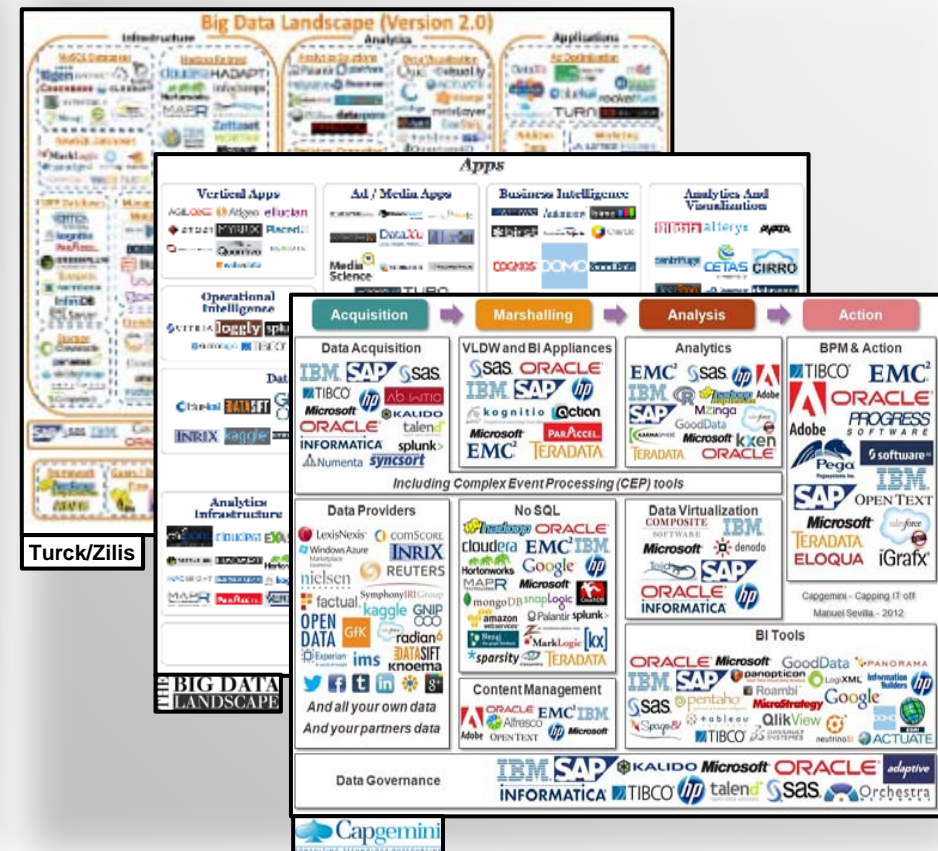**Big Data Applications**

Agenda

1. Open Data & Big Data – Introduction and Definition

2. **Open Data & Big Data – Examples of Big Data Applications**

3. Open Data & Big Data – Selected Challenges

Exemplary companies along the Big Data "value chain" (I/II)

**Big Data
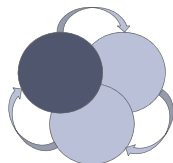"value chain"**
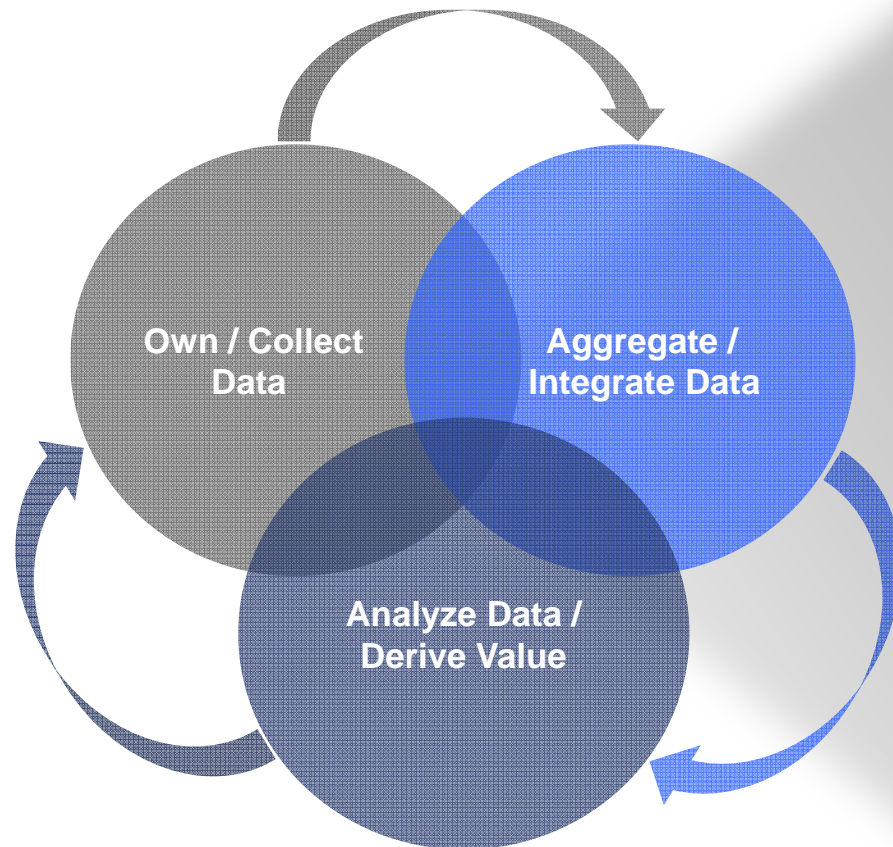
**Players along the Big Data
"value chain"**

Exemplary companies along the Big Data "value chain" (II/II)

Focus on Big Data **Collection**	Focus on Big Data **Aggregation / Integration**	Focus on Big Data **Analytics**

- Primary data collection or control over data access
- ➢ Selling or licensing of data access and data sets
- ➢ E.g. Sale of Twitter data feeds to Gnip

- Provision of the technical infra-structure for data aggregation, data backup, and data management
- ➢ Sale of technical infrastructure service and consulting services
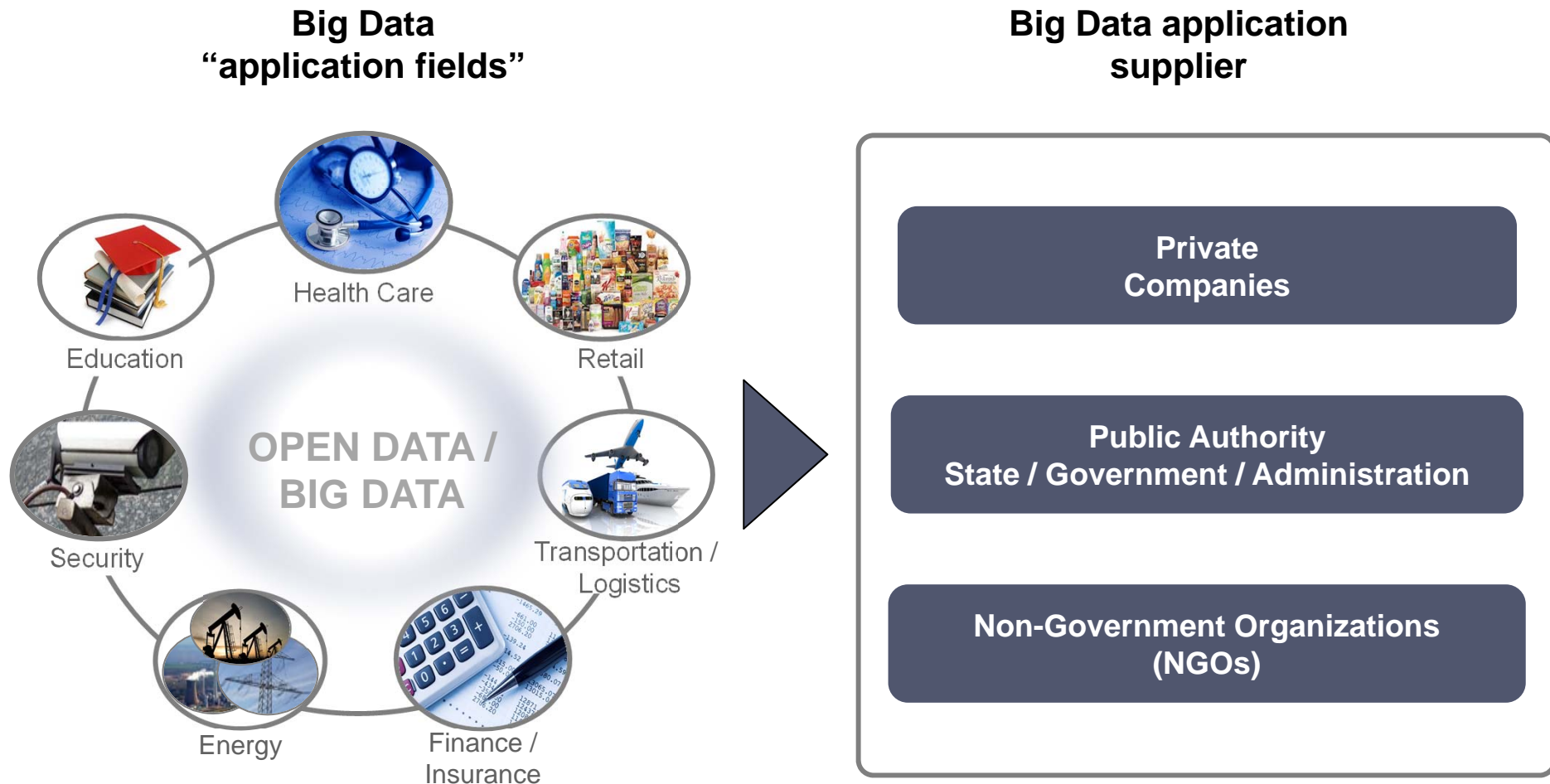- ➢ E.g. Oracle Big Data appliance as an integrated Big Data solution

- Target orientated analysis of Big Data data sets
- ➢ Sale of data analysis and visualization services
- ➢ E.g. Kaggle Connect - crowd-sourced analytics by subject-specific integration of scientists for data analysis
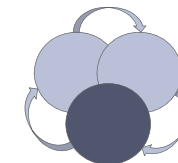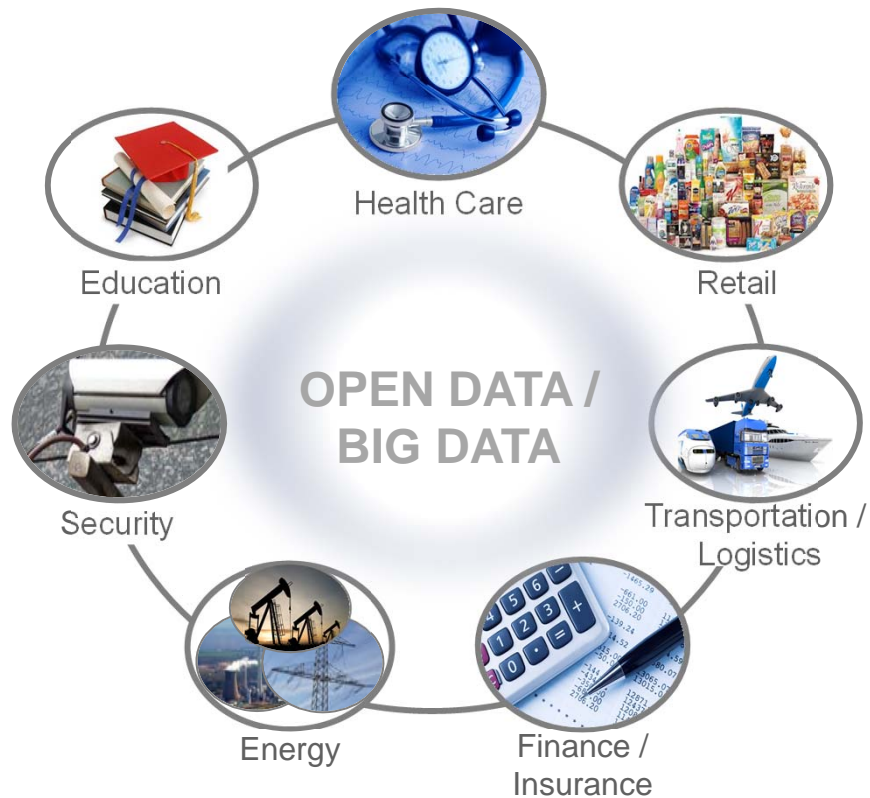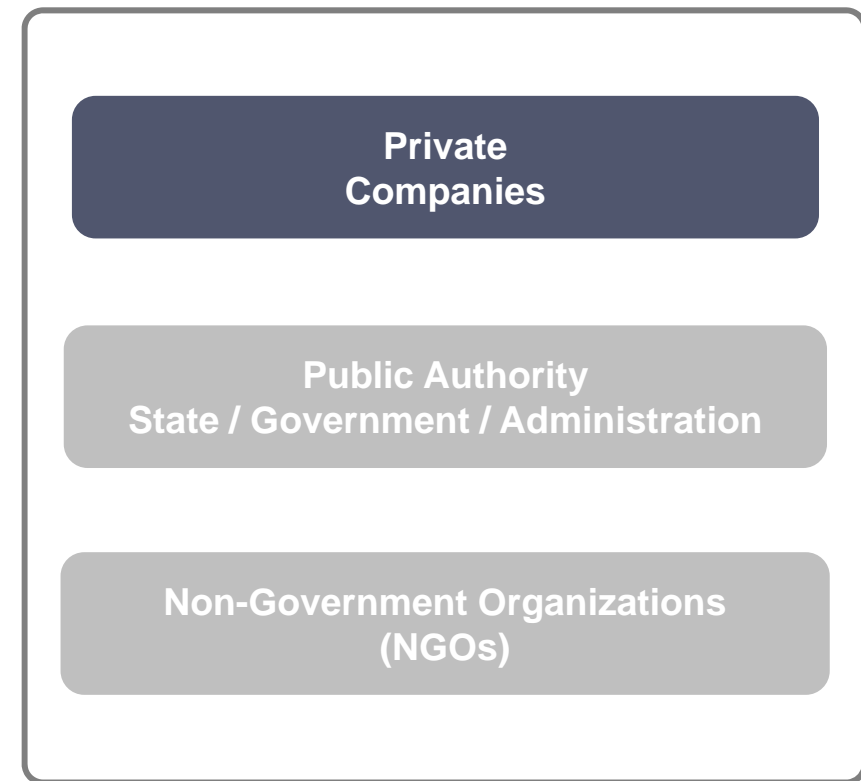
Big Data application fields and its suppliers

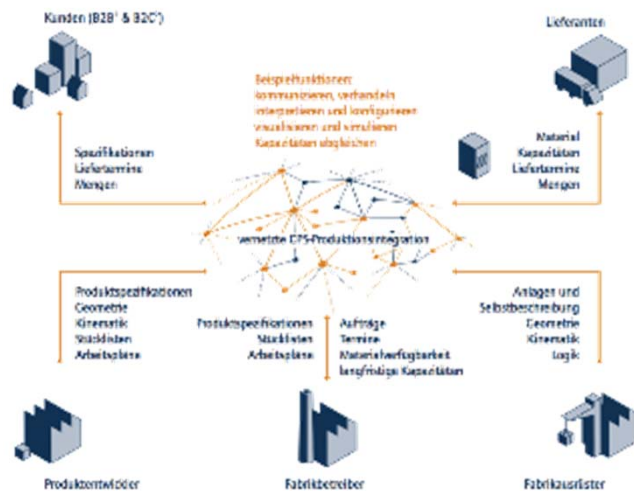**Big Data
"application fields"**

**Big Data application
supplier**

Education

Health Care

Retail

Security

OPEN DATA /
BIG DATA

Transportation /
Logistics

Energy

Finance /
Insurance

**Private
Companies**

**Public Authority
State / Government / Administration**

**Non-Government Organizations
(NGOs)**

Big Data application fields and its suppliers

**Big Data
"application fields"**

**Big Data application
supplier**

Health Care

Education

Retail

OPEN DATA /
BIG DATA

Security

Transportation /
Logistics

Energy

Finance /
Insurance

**Private
Companies**

**Public Authority
State / Government / Administration**

**Non-Government Organizations
(NGOs)**

Big Data adoption in private companies

Focus on **Efficiency** by Big Data

Focus on **Individualization** by Big Data

Focus on **Intelligence/new products** by Big Data

Example: Connected Production

Example: Facebook Newsfeed

Example: Telemedicine

- Direct access to more diverse and more recent data
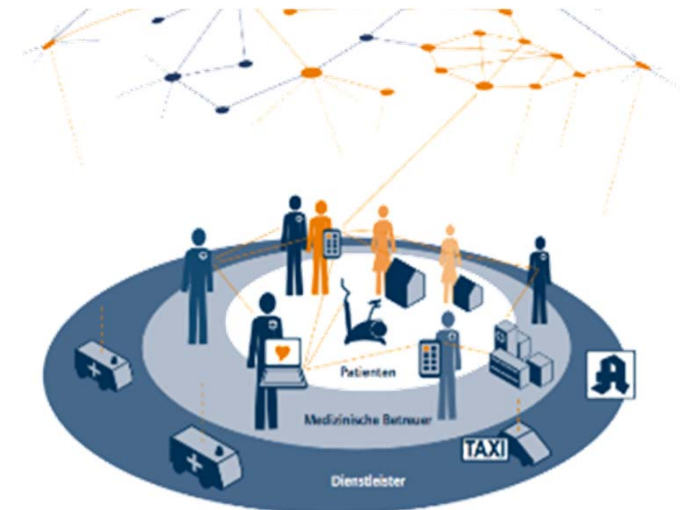- More accurate forecasts
- ➢ Superior performance in production

- Establishment of a privacy justifiable knowledge base about properties and consumers
- ➢ Usage of knowledge base for mass-individualized services

- Products obtain a certain kind of "Integrated intelligence" – e.g. via algorithms
- ➢ Creation of new products and enhancement of existing products by value-added services

Commercial Big Data applications: 23andMe

- 23andMe is a **DNA analysis service** providing information and tools for individuals to learn about and explore their DNA.

- Price: $99 (2007: $999)

- Today: ~ 450,000 customers
  Target: 25 mio. customers
  → Building up a **database of genetic profiles**, the company can look for **disease patterns** and areas of research that **haven't even been considered yet**.

| **Value Proposition**	• With reports on over 240+ health conditions and traits, customers can learn about genetic health risks, carrier status, drug response…and even get information about DNA relatives.
Value Added Architecture	• Usage of Illumina OmniExpress Plus Genotyping BeadChip. • In addition to the variants already included on the chip, 23andMe includes their own customized set of variants relating to conditions and traits.
Revenue Model	• Direct revenues from private customers • Revenues from selling specific data subsets for research purposes

Source: Gutjahr (2013); 23andMe (2013)

Commercial Big Data applications: TomTom Traffic

- **TomTom Traffic** is **a real-time traffic service** that provides information about **jams** on **motorways, major roads and secondary roads**.
- Combination of **information from major traffic authorities** with crowd-sourced **real-time data from over 350 mio. drivers.**
- Data sources:
  - connected GPS devices
  - government loops
  - journalists collecting incident data
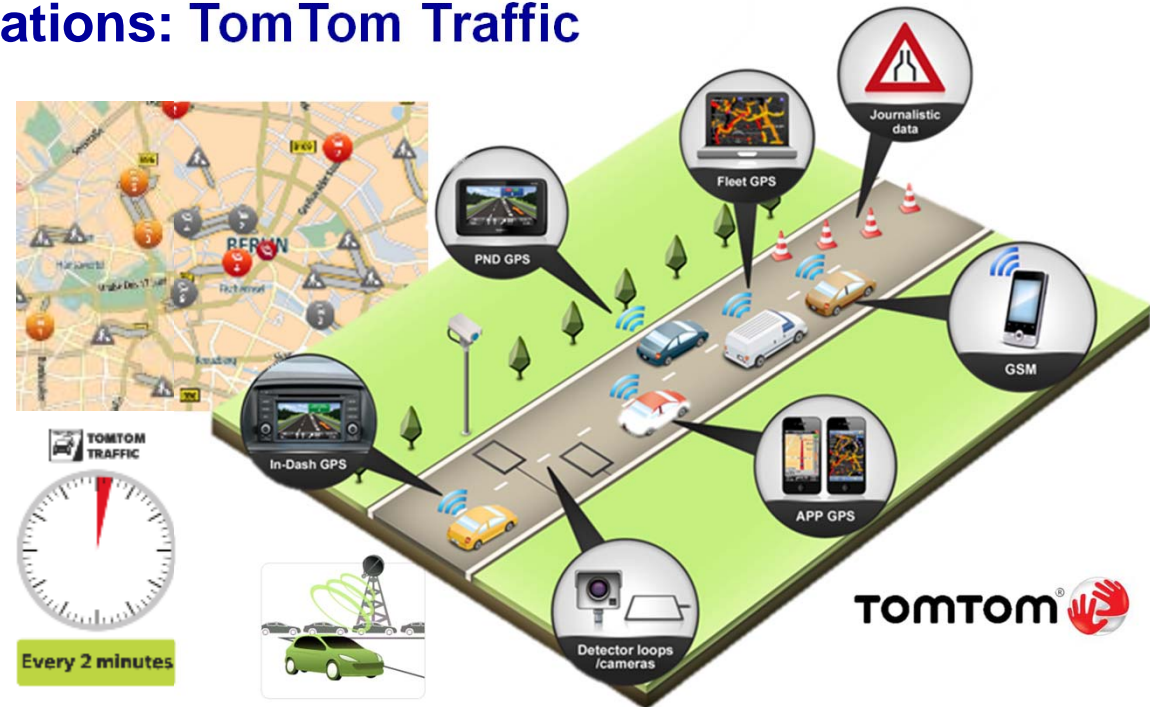
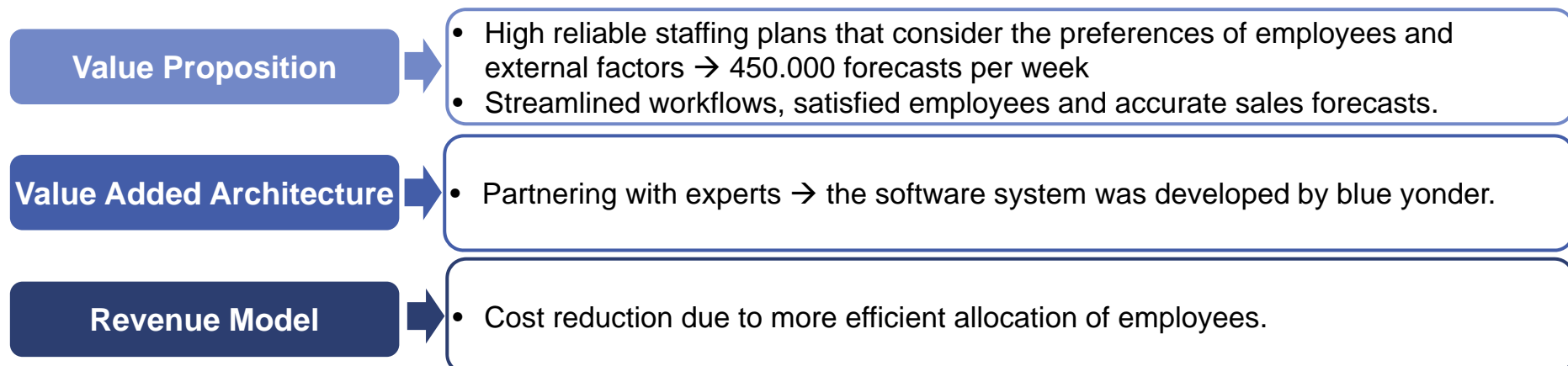| Value Proposition	• Higher efficiency of traffic flow. • Totality of travel times can be reduced.
Value Added Architecture	• Partnering with public authorities and journalists. • Usage of the GPS satellite infrastructure.
Revenue Model	• B2C: Enduser devices incl. data service • B2B: Components especially for cars and traffic management; licenses

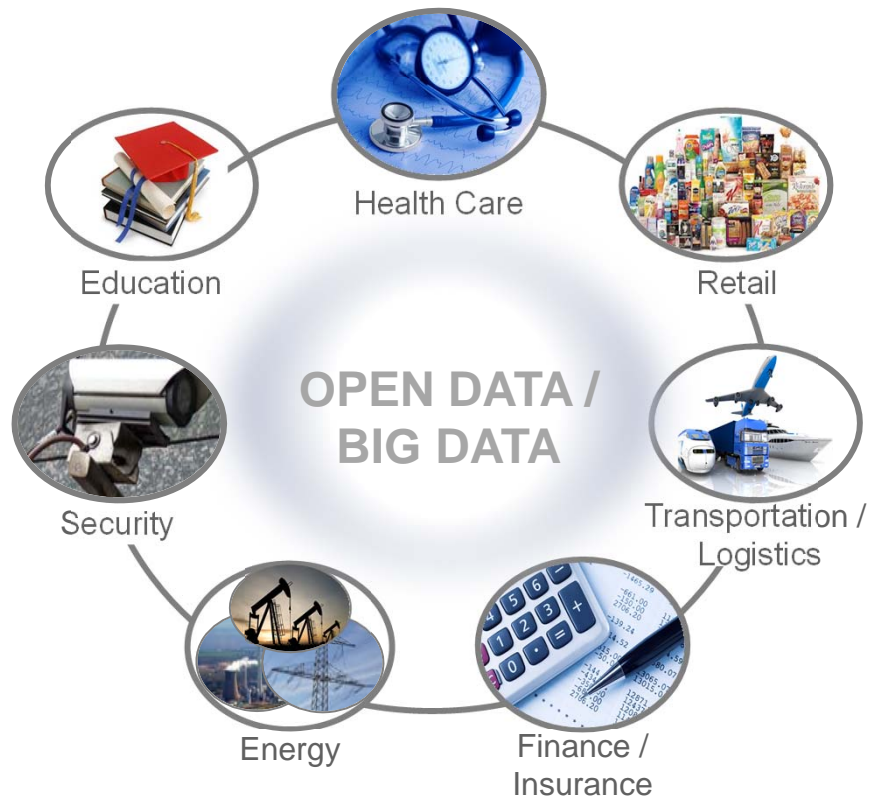Source: Big Data-Startups (2012); TomTom (2013)

24

Commercial Big Data applications: dm

- To accurately **predict** the **daily turnover** of single branches and to **plan** the necessary **number of employees**, dm introduced a **Big Data Software**, developed by blue yonder.

- The solution takes into account the **daily sales** from the past, the **pallet delivery, forecasts** of stock and **individually adjust-able parameters** such as **opening times**.

- In addition, **external data** such as **upcoming market days**, **holidays** in neighbor regions or the **weather forecast** can be integrated.
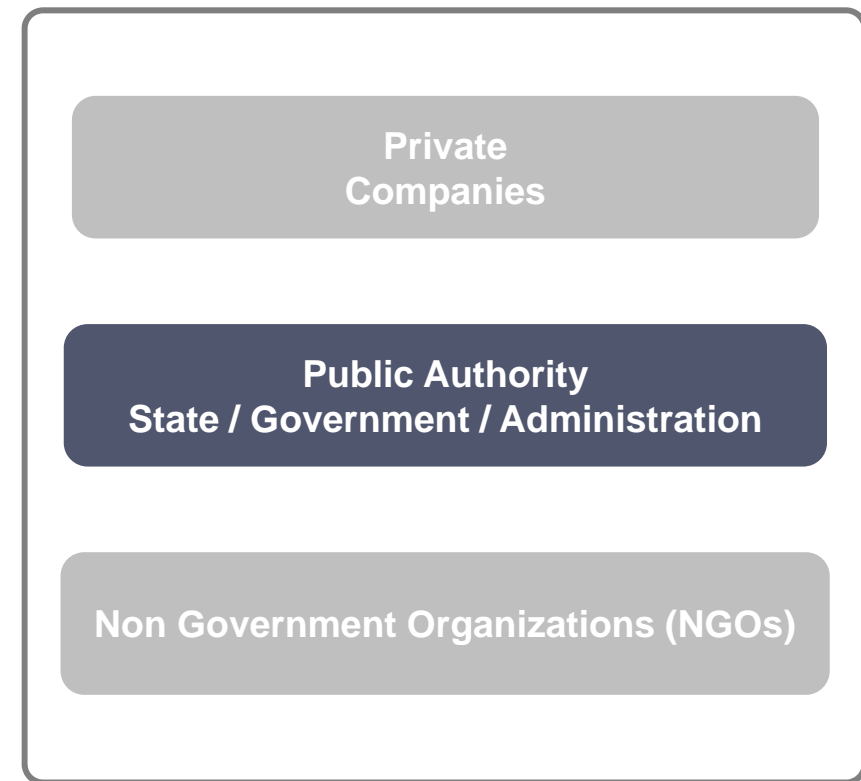
Value Proposition	• High reliable staffing plans that consider the preferences of employees and external factors → 450.000 forecasts per week • Streamlined workflows, satisfied employees and accurate sales forecasts.
Value Added Architecture	• Partnering with experts → the software system was developed by blue yonder.
Revenue Model	• Cost reduction due to more efficient allocation of employees.

Big Data application fields and its suppliers

**Big Data
"application fields"**

**Big Data application
supplier**

OPEN DATA /
BIG DATA

Education

Health Care

Retail

Security

Transportation /
Logistics

Energy

Finance /
Insurance

**Private
Companies**

**Public Authority
State / Government / Administration**

Non Government Organizations (NGOs)

Big Data applications by public authorities: surveillance system in NYC

- **So-called *„Domain Awareness System"* for crime detection and prevention, developed by Microsoft**

- **It monitors and compares data from different sources, including:**
    - 3.000 Video cameras
    - License plate scanners
    - Radiation detectors
    - Diverse data bases

Value Proposition
- Detection and prevention of crime and terrorism → Increase in public safety
- Efficient allocation of security forces

Value Added Architecture
- Partnering with experts → the technical System was developed by Microsoft

Revenue Model
- Public financing; cost reduction due to a higher efficiency in crime prevention
- If the system is licensed out to other cities, New York City will get 30% of the profits

Source: New York Daily News (2012); Zeit (2012)

Big Data applications by public authorities: data.gov

- **In 2009, the US government launched the website „data.gov"**

- **Public access to machine readable datasets generated by the Executive Branch of the Federal Government:**
  - 91,100 datasets
  - 349 citizen-developed apps
  - 137 mobile apps
  - 175 agencies and subagencies
  - 87 galleries
  - 409 Government APIs

|	
Value Proposition	• Improving access to federal data and expanding the creative use of those data • Encouraging innovative ideas (e.g. web applications) • Making government more transparent → strengthen Nation's democracy
Value Added Architecture	• Public participation & collaboration → open access to datasets that can be used to build applications, conduct analyses, and perform research • Processing user's feedback and ideas to improve or enhance the data supply
Revenue Model	• Financed by the US Electronic Government Fund

Source: Data.gov (2013)

Big Data applications by public authorities: United Nations Global Pulse

- Investigation **whether** and **how social media** and other online user-generated content could enrich **understanding** of the effect of **changing employment** conditions.

- **Goal:** Comparison of the **qualitative information** offered by **social media** with **unemployment figures**

- Usage of SAS Social Media Analytics and SAS Text Analytics to dig into **data from 500,000 blogs, forums, and news sites in the US and Ireland**
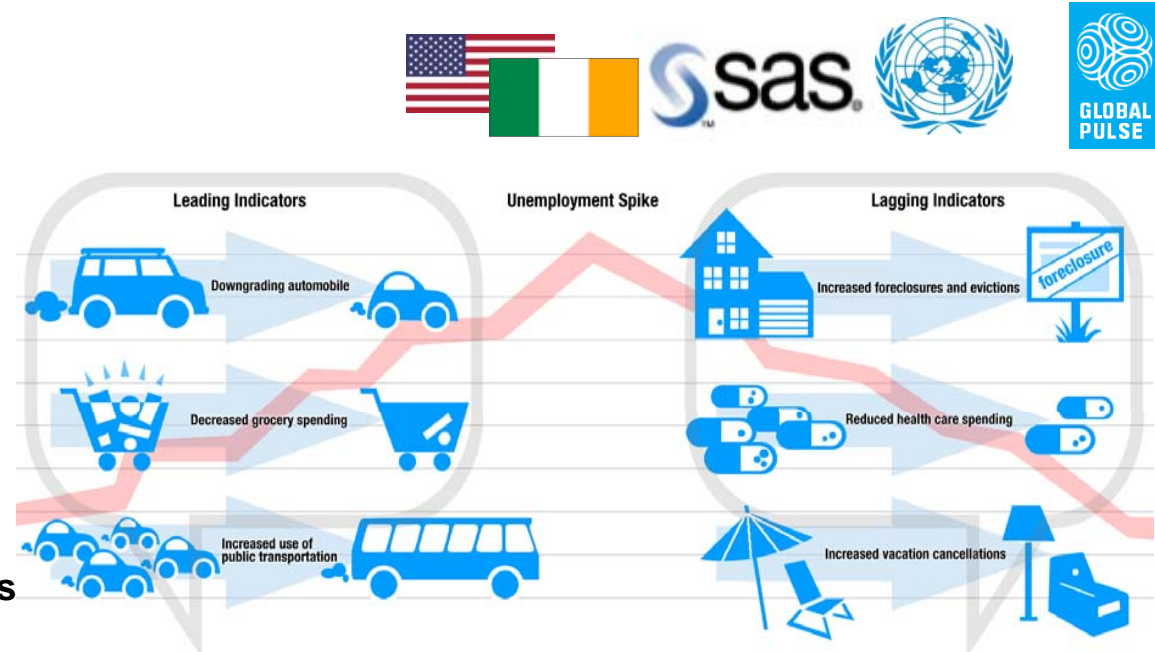
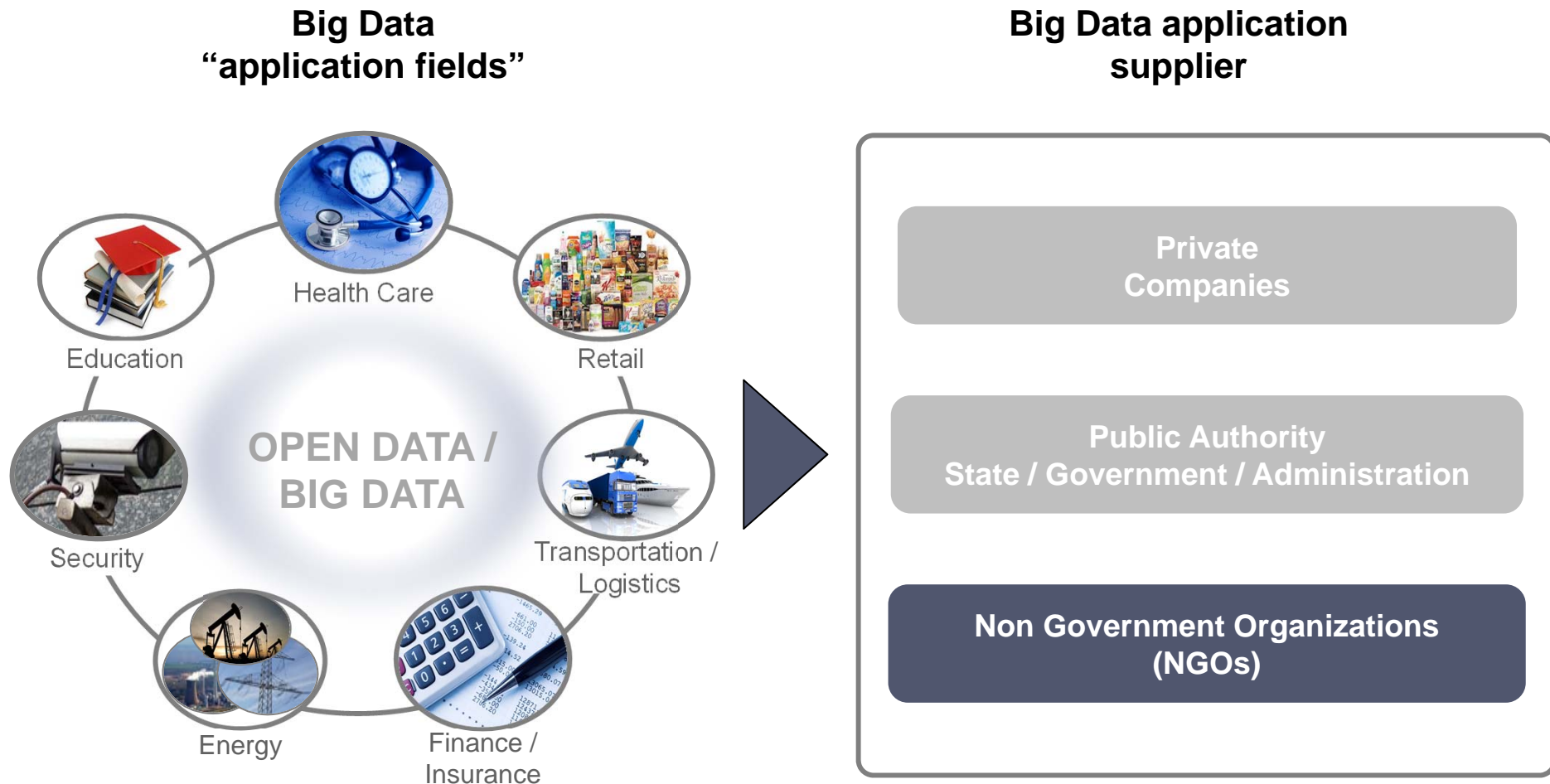| **Value Proposition** | • Insight into how people are coping with crises
• Forecasting changes in the employment rate before official statistics are provided
• Identification of correlations relevant for the labor market |

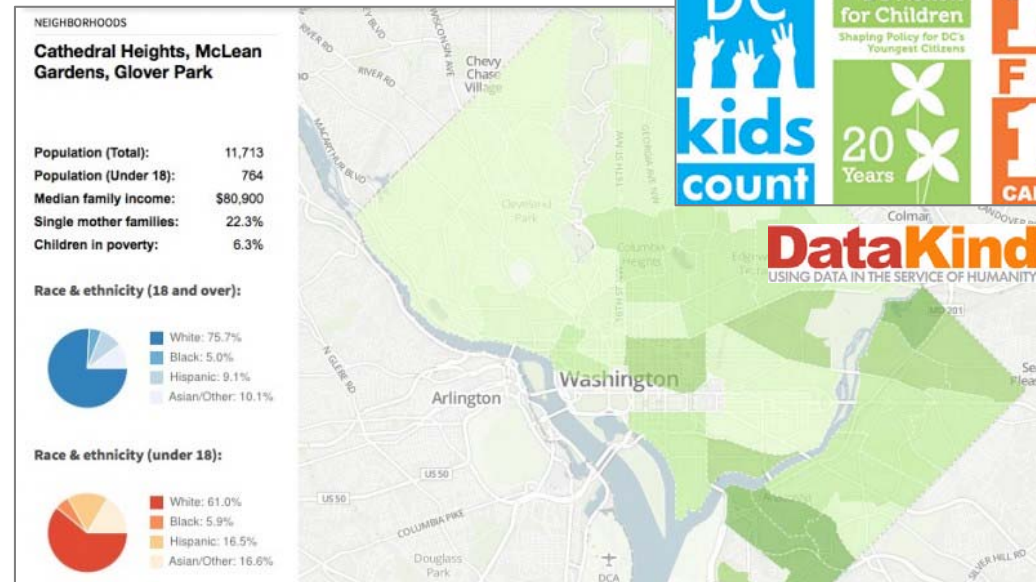| **Value Added Architecture** | • Partnering with experts → the technical System was developed by SAS |

| **Revenue Model** | • Funded by voluntary contributions of single UN-member states and private organizations |

Source: SAS (2013); United Nations Global Pulse (2011)

Big Data application fields and its suppliers

**Big Data
"application fields"**

**Big Data application
supplier**

Health Care

Education

Retail

OPEN DATA /
BIG DATA

Security

Transportation /
Logistics

Energy

Finance /
Insurance

**Private
Companies**

**Public Authority
State / Government / Administration**

**Non Government Organizations
(NGOs)**

Big Data applications by NGOs: DC Action for Children

- **DC Action for Children** is a **data-driven NGO**, based in Washington, D.C.

- The organization **helps children** who are most in need of health, education, safety and financial well-being.

- DC Action for Kids collects **data** in form **of official statistics and own surveys.**

- **2012: Cooperation** with **DataKind** → They created **interactive city maps,** in which the **living conditions** in various **districts** along **different dimensions** are visible.

NEIGHBORHOODS

Cathedral Heights, McLean Gardens, Glover Park

Population (Total):	11,713
Population (Under 18):	764
Median family income:	$80,900
Single mother families:	22.3%
Children in poverty:	6.3%

Race & ethnicity (18 and over):
- White: 75.7%
- Black: 5.0%
- Hispanic: 9.1%
- Asian/Other: 10.1%

Race & ethnicity (under 18):
- White: 61.0%
- Black: 5.9%
- Hispanic: 16.5%
- Asian/Other: 16.6%

|	
Value Proposition	• Creation of an interactive databook, replacing a traditional PDF factsheet, collecting crunched data on the indicators that have an impact on poverty • Visualization of living conditions in single districts → efficient allocation of assistance
Value Added Architecture	• Partnering with experts → DC Action for children has the local knowledge on where to find data; DataKind has the competence to build visualizations of traditional data
Revenue Model	• DC Action for Children is funded by several nonprofits, foundations and corporations

31

Additional Big Data applications: IBM Watson

- In 1997, a large-capacity computer beat the chess champion Garry Kasparov the first time (IBM's Deep Blue, that was able to calculate more than 200 possible moves per second)

- Watson relies on understanding natural human language, analyzing the words and context, processing this information quickly and, subsequently, providing the answers to questions in natural language

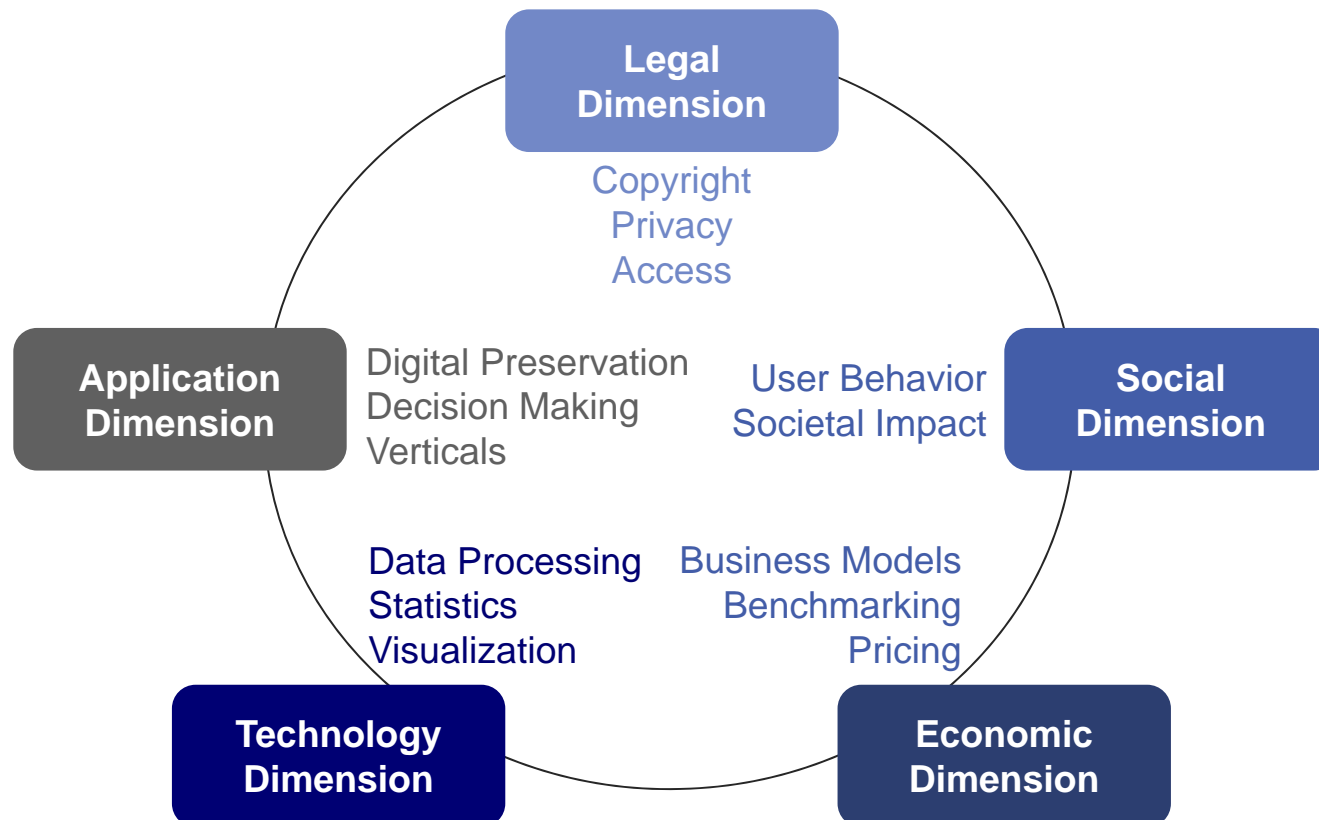- In 2011, Watson beat two Jeopardy!-Champions
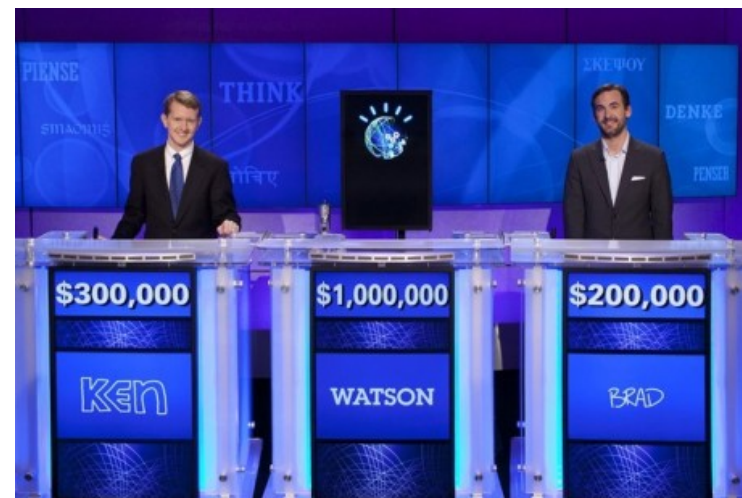
Additional Big Data applications: "Handshake" - data marketplace

Data Marketplace
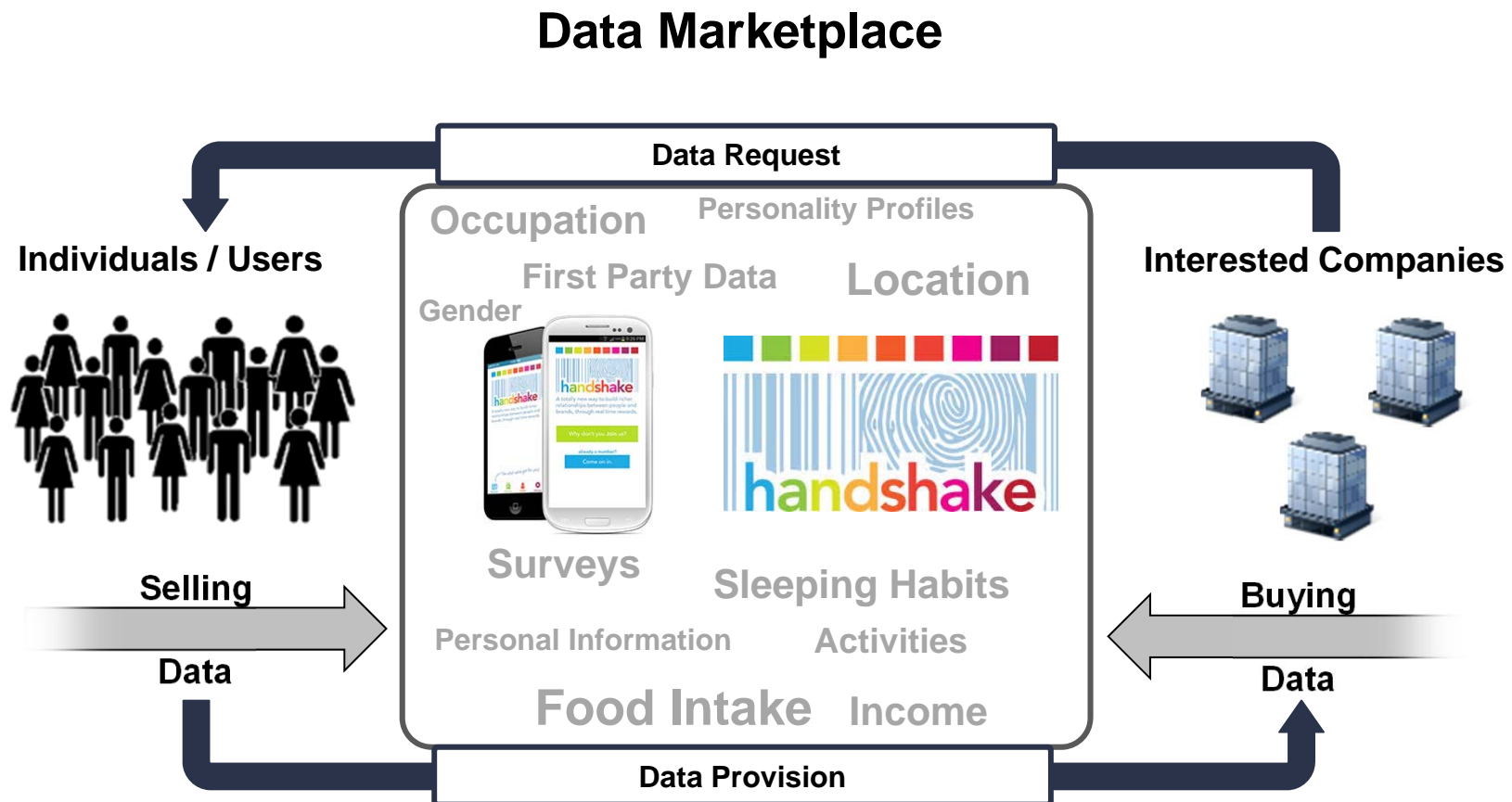
Agenda

1. Open Data & Big Data – Introduction and Definition

2. Open Data & Big Data – Examples in Private Life and Business

3. **Open Data & Big Data – Selected Challenges**

LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

RESEARCH CENTER FOR INFORMATION,
ORGANIZATION AND MANAGEMENT
PROF. DR. DRES. H.C. ARNOLD PICOT

Challenges in all dimensions: „Big Data Space"

Selected challenges

**Data-
level**

- What data should be stored?
- Who has access to the data sets?
- Who can use data?
- …

**Technology-
level**

- Can the required analyzes be carried out cost-effectively?
- Are the in future required IT skills of employees and companies sufficiently available?
- …

**Usage-
level**

- What data are (or possibly will be) individual-related/personal?
- How can social acceptance be achieved?
- …

Selected legal challenges

- **Big Data as a diverse, meaningful asset**
- **No guaranteed ownership of data by civil law – protection gap in the long run?**
- **Right of the database vendors**
    - Data collection, not single data sets
    - Ownership-similar status limited to 15 years
    - High demand for contractual arrangements
- **Data privacy is an important factor, but**
    - In case of Big Data, often irrelevant
    - Prohibition with reservation of authorization increasingly questionable
- **Issues regarding the competition law:**
  → Is there a need to release privately collected data ("essential facilities")?
- **High potential for innovation by open data**

| **From "Data" to "Big Data"** |
| Conflicts are likely to increase: So far "small problems" will turn into "big problems" |

Thank you for attention!

FROM OPEN DATA TO BIG DATA

RESEARCH CENTER FOR INFORMATION,
ORGANIZATION AND MANAGEMENT
PROF. DR. DRES. H.C. ARNOLD PICOT

Sources (1/4)

23andMe (2013): How it works; 11.11.2013; URL: https://www.23andme.com/howitworks/

Acatech (2012): Integrierte Forschungsagenda Cyber-Physical Systems; URL: http://www.acatech.de/ fileadmin/user_upload/Baumstruktur_nach_Website/Acatech/root/de/Material_fuer_Sonderseiten/Cyber-Physical-Systems/acatech_STUDIE_agendaCPS_Web_20120312_superfinal.pdf

Beyer, M. & Laney, D. (2012): The Importance of 'Big Data': A Definition, Gartner Research Report

Big Data-Startups (2012): TomTom and Big Data; 11.11.2013; URL: http://www.bigdata-startups.com/BigData-startup/tomtom-big-data/

BITKOM (2012): Big Data im Praxiseinsatz – Szenarien, Beispiele, Effekte

blue yonder (2013): Effiziente Mitarbeitereinsatzplanung dank exakter Prognosen; 11.11.2013; URL http://www.blue-yonder.com/dm-drogerie-markt.html

Boyd, D. & Crawford, K. (2012): Critical Questions for Big Data, Information, Communication & Society, Vol. 15(5), pp. 662-679.

Capgemini (2012): Big Data vendors and technologies, the list!; 11.11.2013; URL: http://www.capgemini.com/ blog/capping-it-off/2012/09/big-data-vendors-and-technologies-the-list

CIO (2012): Anwenderbeispiele für Big Data; 11.11.2013; URL: http://www.cio.de/index.cfm?pid=249&pk= 11218&fk=684837

DataKind (2012): Mapping Poverty to Beat It; 11.11.2013; URL: http://www.datakind.org/mapping-poverty-to-beat-it/

Data.gov (2013): About data.gov; 11.11.2013; URL: http://www.data.gov/about

Sources (2/4)

DC Action For Children (2013): About Us; 11.11.2013; URL: http://www.dcactionforchildren.org/content/about

dm (2013): Pressemitteilungen; 11.11.2013; URL: https://www.dm.de/de_homepage/presse/
pressemitteilungen

Duisberg, A. (2012): Wem gehören die Daten und wer hat außerdem Rechte daran?, in: Big Data wird neues
Wissen, Fachvortrag der am 24. Mai 2012 in München abgehaltenen Fachkonferenz; 11.11.2013; URL:
http://www.muenchner-kreis.de/pdfs/BigData/Duisberg.pdf

Eberspächer, J. & Wohlmuth, O. (2012): Big Data wird neues Wissen, Vorträge der am 24. Mai 2012 in
München abgehaltenen Fachkonferenz; URL: http://www.muenchner-kreis.de/?id=339

Economist (2010): Data, data everywhere; 25.02.2010; URL: http://www.economist.com/node/15557443

Fraunhofer IAIS (2012): Big Data – Vorsprung durch Wissen; URL: http://www.iais.fraunhofer.de/fileadmin/
user_upload/Abteilungen/KD/pdfs/FraunhoferIAIS_Big-Data_2012-12-10.pdf

Gutjahr, R. (2013): Sie haben ein erhöhtes Risiko für Prostatakrebs - Interview mit Catherine Afarian,
Sprecherin des Unternehmens 23andMe, in: Frankfurter Allgemeine Zeitung, No. 259/45, p. 27.

Handshake (2013): What is Handshake?; 11.11.2013; URL: http://handshake.uk.com/hs/what-is-
handshake.html

IDC/Dell (2012): Big Business dank Big Data? Neue Wege des Datenhandlings und der Datenanalyse in
Deutschland 2012; 19.10.2012; URL: http://www.idc.de/press/presse_idc-studie_big_data2012.jsp

Markl, V. (2012): Research, Innovation and Teaching in Big Data Analytics, Fachvortrag Münchner Kreis,
München im Mai 2012

Sources (3/4)

McKinsey Global Institute (2013): Open data: Unlocking innovation and performance with liquid information, Report October 2013

New York Daily News (2012): NYPD unveils new $40 million super computer system that uses data from network of cameras, license plate readers and crime reports; 11.11.2013; http://www.nydailynews.com/new-york/nypd-unveils-new-40-million-super-computer-system-data-network-cameras-license-plate-readers-crime-reports-article-1.1132135

Picot, A. & Reichwald, R. & Wigand, R.T. (2008): Information, Organization and Management, Springer Verlag, Berlin

SAS (2013): UN Global Pulse honored for research using SAS® social media, text analytics; 11.11.2013; http://www.sas.com/news/preleases/analytics-computerworld-award.html

Spiegel (2013): Leben nach Zahlen, 13.05.2013, No. 20.

Techcrunch (2013): Handshake Is A Personal Data Marketplace Where Users Get Paid To Sell Their Own Data; 11.11.2013; URL: http://techcrunch.com/2013/09/02/handshake/

The Big Data Landscape (2013): The Big Data Landscape; 11.11.2013; URL: http://www.bigdatalandscape.com

TomTom (2013): TomTom Traffic; 11.11.2013; URL: http://www.tomtom.com/de_de/services/live/hd-traffic/

Trendreport Betterplace Lab (2013): Big Data for Good; 11.11.2013; URL: http://trendreport.betterplace-lab.org/trend/big-data

Turck, M. & Zilis, S. (2012): A chart of the big data ecosystem, take 2; 11.11.2013; URL: http://mattturck.com/2012/10/15/a-chart-of-the-big-data-ecosystem-take-2/

Sources (4/4)

United Nations Global Pulse (2011): Unemployment Through the Lens of Social Media; 11.11.2013; URL: http://www.unglobalpulse.org/projects/can-social-media-mining-add-depth-unemployment-statistics

Zeit (2012): Das menschliche Gesicht von "Big Data"; 11.11.2013; URL: http://www.zeit.de/digital/2012-12/fs-human-face-big-data