

# **Harnessing Big Data to Benefit the Transit Agency: Real Time, Fare Media, Ridership ...**

**APTA Legal 2018**

Jenifer Ross-Amato  
Deputy General Counsel



# Data may be the agency's most valuable asset!

- *Data can help transit agencies:*

- Increase market share
- Improve transit services
- Plan for better projects
- Become more fiscally sustainable
- Build multi-modal, smart cities

- *Data can hurt transit agencies:*

- Give critics ammunition
- Help competitors
- Add liability exposure
- Increase costs of litigation
- Overwhelm with open records requests





## 40 ZETTABYTES

[ 43 TRILLION GIGABYTES ]  
of data will be created by 2020, an increase of 300 times from 2005

6 BILLION PEOPLE  
have cell phones

WORLD POPULATION: 7 BILLION

## Volume SCALE OF DATA

## It's estimated that 2.5 QUINTILLION BYTES

[ 2.3 TRILLION GIGABYTES ]  
of data are created each day

Most companies in the U.S. have at least  
**100 TERABYTES**  
[ 100,000 GIGABYTES ]  
of data stored

# The FOUR V's of Big Data

From traffic patterns and music downloads to web history and medical records, data is recorded, stored, and analyzed to enable the technology and services that the world relies on every day. But what exactly is big data, and how can these massive amounts of data be used?

As a leader in the sector, IBM data scientists break big data into four dimensions: **Volume, Velocity, Variety and Veracity**

Depending on the industry and organization, big data encompasses information from multiple internal and external sources such as transactions, social media, enterprise content, sensors and mobile devices. Companies can leverage data to adapt their products and services to better meet customer needs, optimize operations and infrastructure, and find new sources of revenue.

By 2015  
**4.4 MILLION IT JOBS**  
will be created globally to support big data, with 1.9 million in the United States

As of 2011, the global size of data in healthcare was estimated to be

**150 EXABYTES**  
[ 161 BILLION GIGABYTES ]



**30 BILLION  
PIECES OF CONTENT**  
are shared on Facebook every month



## Variety DIFFERENT FORMS OF DATA

By 2014, it's anticipated there will be

**420 MILLION  
WEARABLE, WIRELESS  
HEALTH MONITORS**

**4 BILLION+  
HOURS OF VIDEO**  
are watched on  
YouTube each month



**400 MILLION TWEETS**  
are sent per day by about 200 million monthly active users



The New York Stock Exchange captures  
**1 TB OF TRADE  
INFORMATION**  
during each trading session



## Velocity ANALYSIS OF STREAMING DATA

Modern cars have close to  
**100 SENSORS**  
that monitor items such as fuel level and tire pressure



By 2016, it is projected there will be  
**18.9 BILLION  
NETWORK  
CONNECTIONS**  
— almost 2.5 connections per person on earth



**1 IN 3 BUSINESS  
LEADERS**  
don't trust the information they use to make decisions

**27% OF  
RESPONDENTS**  
in one survey were unsure of how much of their data was inaccurate

## Veracity UNCERTAINTY OF DATA

Poor data quality costs the US economy around  
**\$3.1 TRILLION A YEAR**

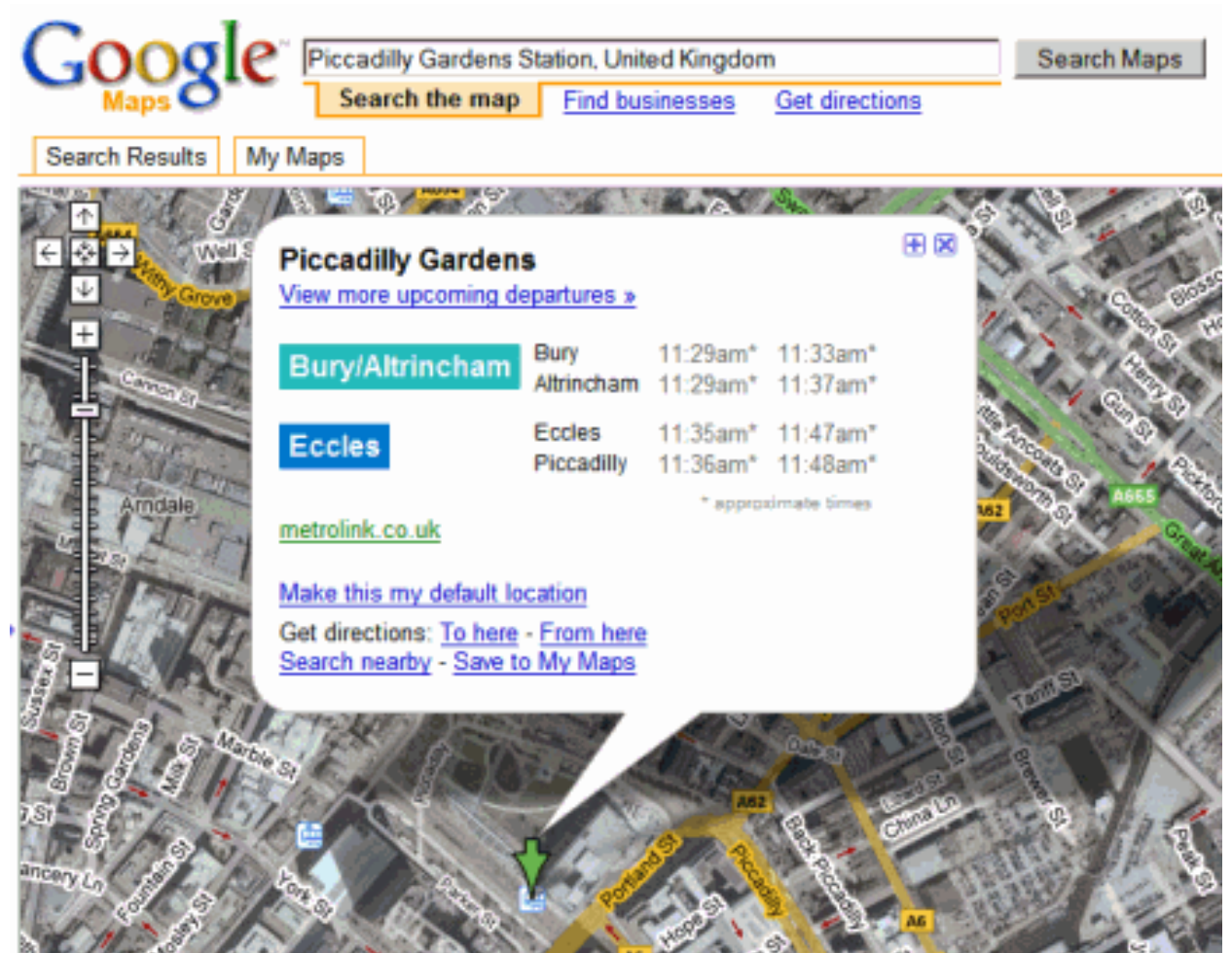


Sources: McKinsey Global Institute, Twitter, Cisco, Gartner, EMC, SAS, IBM, MEPTec, QAS

IBM

# Where do transit agencies collect data?

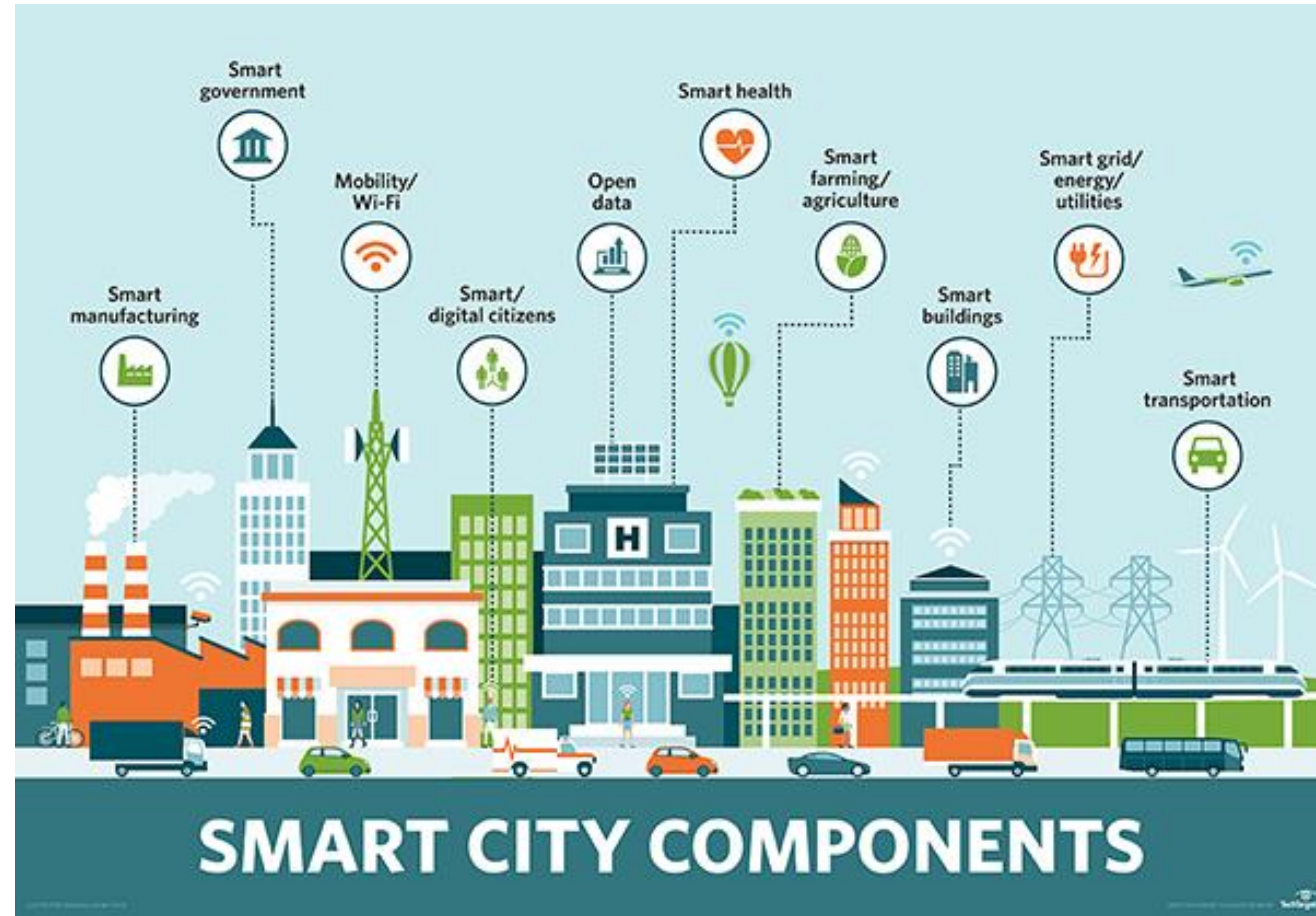
- GIS data
- Smart cards; fare machines
- Mobile ticketing
- Ridership (i.e. boardings per station)
- Customer surveys
- Social media
- Accidents
- Safety incidents
- Video – public spaces, vehicles
- Wifi usage





# Who wants data collected by transit agencies?

- Cities
- Advocacy groups
- School districts
- Competitor transit providers
- Regulators
- Law enforcement



# How is that data stored and accessed?

- Is the data in the cloud or on a local server?
- Who owns the data?
- Is the data in the United States?
- Can you easily access the data and use it for what you need - at a reasonable cost?



# How are you going to protect the data?

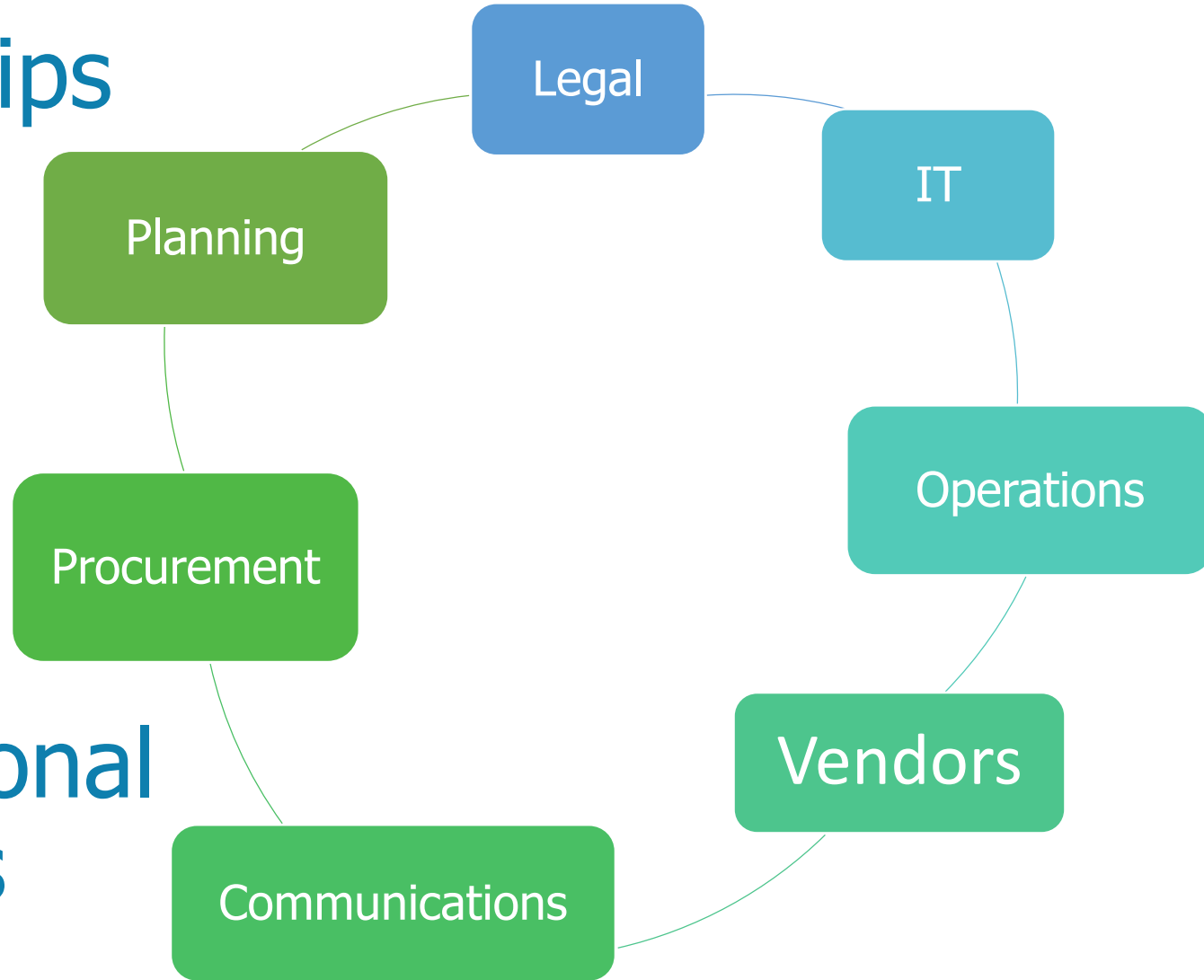
- Who is responsible if the data is lost or stolen?
- Is the data confidential – PII, PCI, PHI?
- Are you legally responsible to a third party (i.e. bank, credit card companies) if the data is compromised?
- If you share the data, how do you ensure its integrity?





# Partnerships

# Organizational Challenges



- Data sharing agreements
- Vendor agreements
- Mega IT projects
- Smart city strategic planning
- Grants management
- Policies & Procedures

# The RTD Denver Story

- Neighborhood and student smart card data sharing & privacy concerns
- Technology procurement project
- Ridership data requests



# Key Legal Provisions

- Form and content of data – aggregate, anonymized
- Limitation of liability
- Data breach notification
- Demonstration of legal compliance versus audit rights
- License and maintenance agreements
- Ability to access and extract data – key costs
- Approved uses of data and sharing with other parties
- No ongoing obligations; limited terms
- Integrity of the data
- Identifying and protecting CSSI