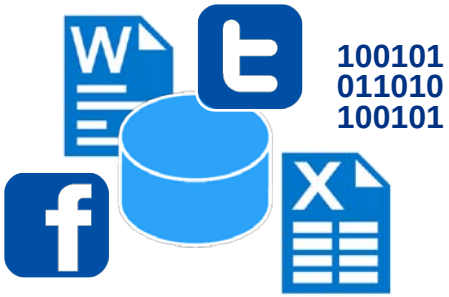
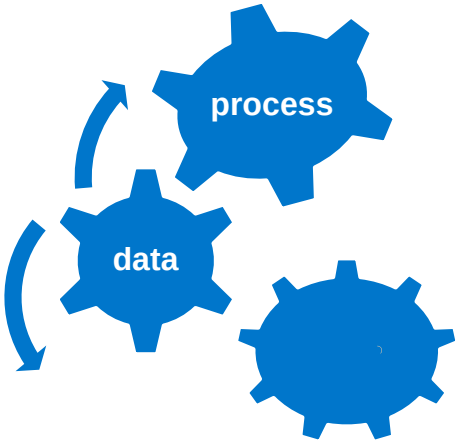

Big Data
Discussion

What is Big Data?

Gaining Instant Insight from data in the format you need for better decisions





Dashboard

Ad hoc analysis

Reports

Gather All Relevant Big Data

Insight to Action

Data diversity formats make it difficult to analyze

Pressure to gain insight quickly from data

Need to determine the right action from many valuable insights

New Challenges & Big Data Require A Different Approach

Leaders Are Breaking The Traditional Information Management Model

Traditional Approach

Business Users
Determine what question to ask

IT
Structures the data to answer that question

Structured & Repeatable Analytics

- Query Based -- Questions Drive Data
- Customer Surveys & Focus Groups
- Monthly, Weekly, Daily
- Data At Rest

VS.

Big Data Approach

IT
Delivers a platform to enable creative discovery

Business
Explores what questions could be asked

Iterative & Exploratory Analytics

- Autonomic -- Insight Drives Answers
- Customer Sentiment
- Persistent & Ad Hoc
- Data In Motion & at rest

An Explosion of Data

By 2016, annual Internet traffic will reach **1.3 Zettabytes**

(1 ZB = 1,000,000,000,000,000,000,000 bytes)

Facebook processes **500+ Terabytes** of data daily

150 Exabytes

global size of “Big Data” in Healthcare, growing between 1.2 and 2.4 EX / year

Hadron Collider at CERN generates **40 Terabytes** of usable data / day

Google processes **> 24 Petabytes** of data in a single day

Twitter processes **12 Terabytes** of data daily

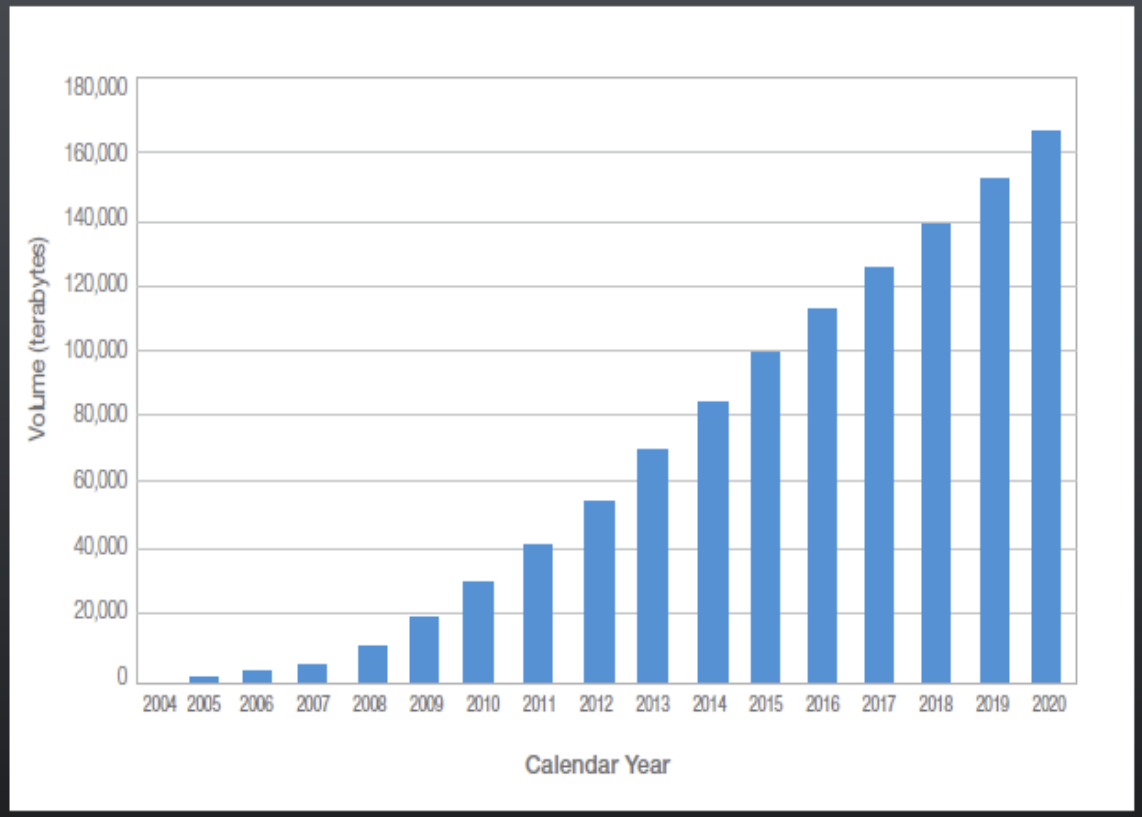
AT&T transfers about **30 Petabytes** of data through its network daily

For every session, NY Stock Exchange captures **1 Terabyte** of trade information

Exponential Data Growth

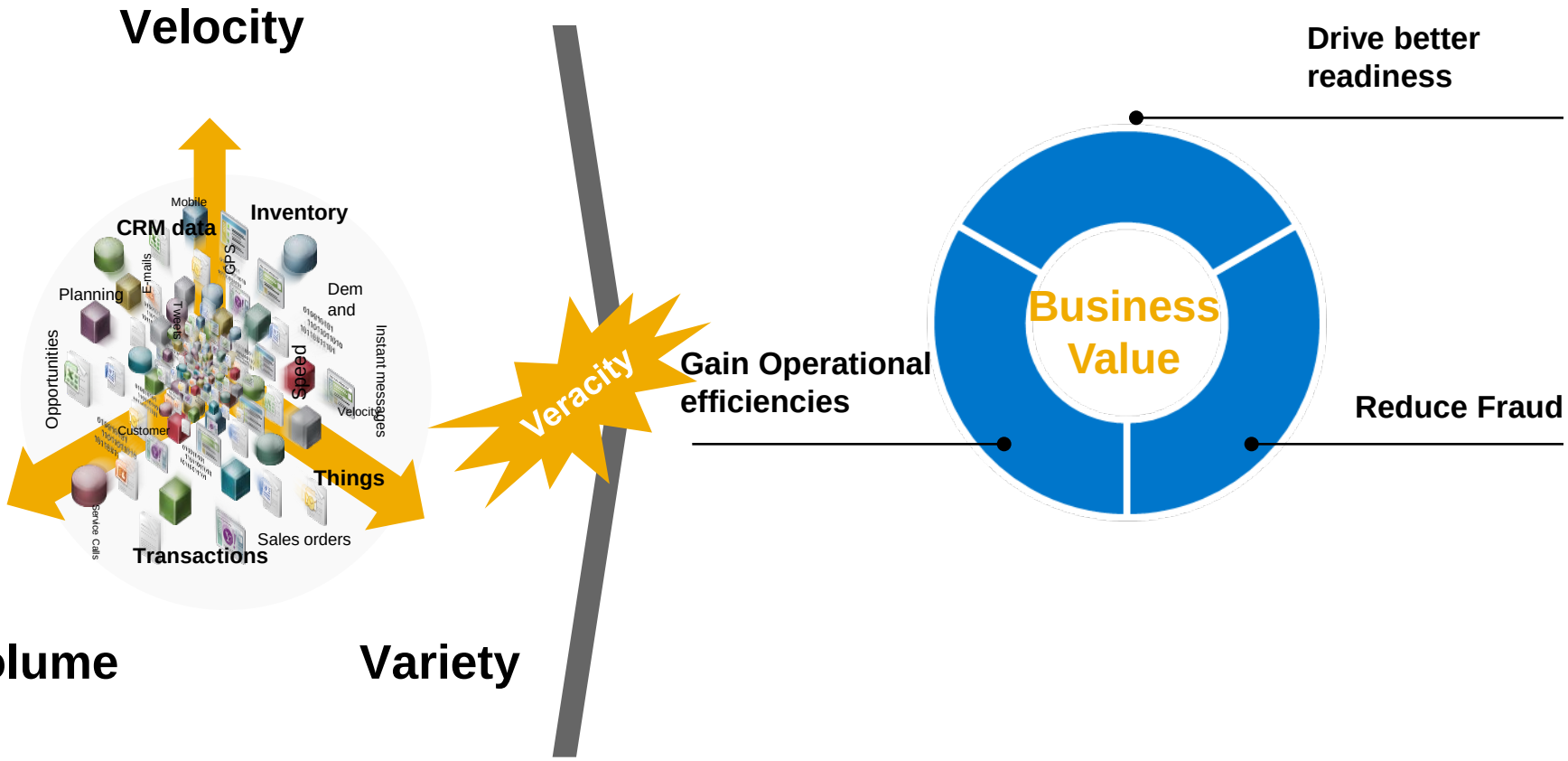
The National Oceanic and Atmospheric Administration's (NOAA) data management system predicts that the agency's amount of archived data will grow to more than 160,000 terabytes (TB) by 2020, due primarily to huge amounts of data being collected by remote sensing of the atmosphere, oceans, land and space.

Exponential data growth isn't limited to large federal agencies like NOAA. For example, the data storage capacity requirements of Clackamas County, Ore., increased from 4 TB in 2005 to around 60 TB in 2010.[19]

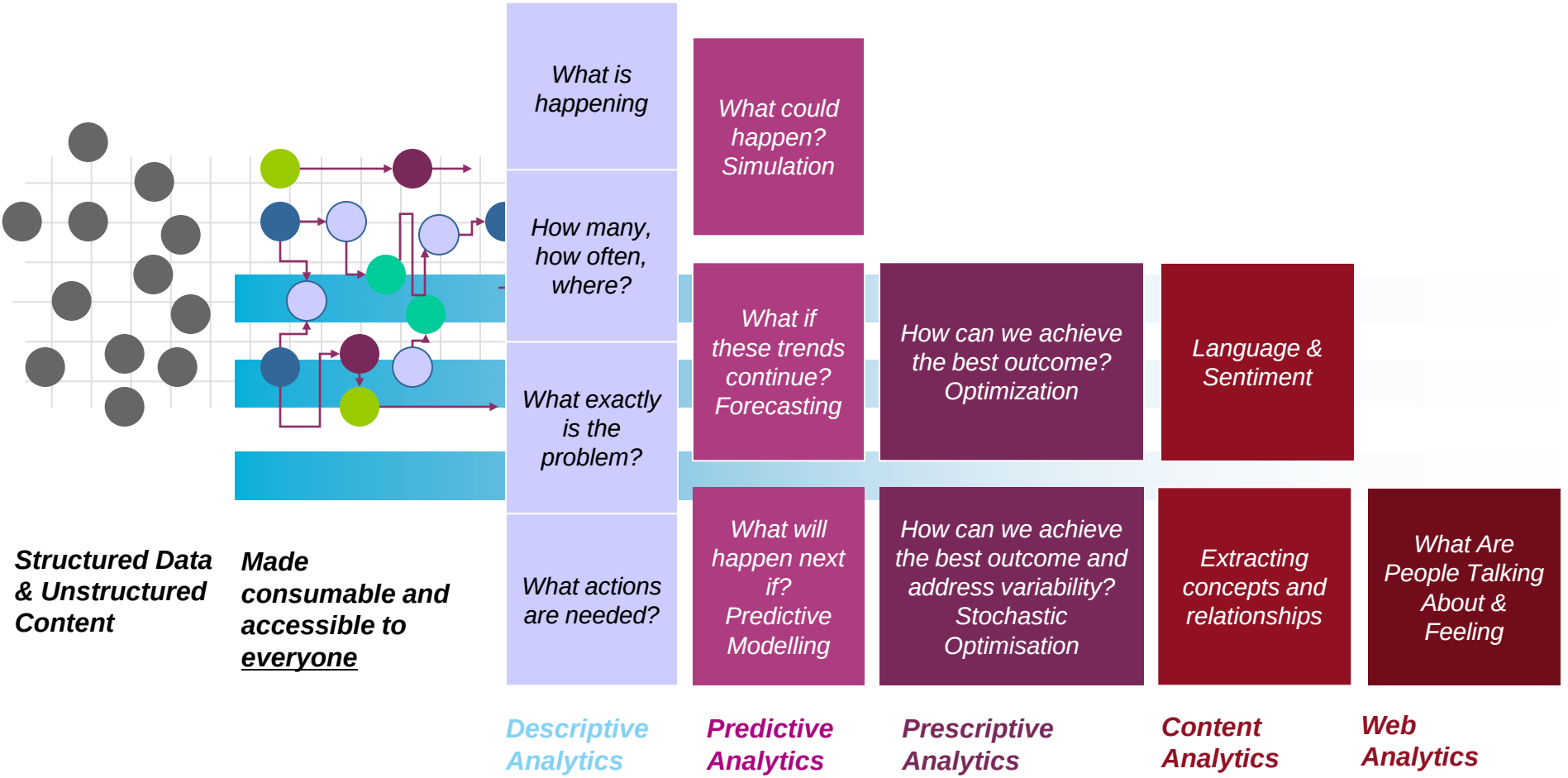


Source: http://celebrating200years.noaa.gov/visions/data_mgmt/side1_class.html

3V's... Versus 5 V's



Big Data Enables Different Kinds of Analytics



Challenge in Government:
Data is Diverse, Structured, Unstructured and Growing

Traditional Data Sources

- **Claims**

- **Tax Files**

- **Crime Data**

- **External Agencies**

- **Beneficiary**

- **Assets**

- **Case Information**

- **Logistics**

- **Adverse Event**

- **Finance**

Non-Traditional Data Sources

- **Sensors**

- **Imagery**

- **Video**

- **Internet Traffic**

- **Social Media**

What are the government problems that require insight with Big Data?

The Use Cases Are Endless

Optimize evacuation routes • Model economic growth • Real-time situational awareness • Boost cyber security • Expedite intelligence gathering • Rapid funds availability • Improve spend analysis • Enhance predictive maintenance • Increase asset availability • Better predict system failures • Sense & respond in real-time • Traffic impact • Stop Improper payments before they are paid• Model economic impacts • Predict weather impact • Understand Citizen Sentiment • Optimize use of excess energy • Pinpoint environmental risks • Understand crime trends •

Imagine the Possibilities of Harnessing Government Data Resources

Surveillance system analyzes & classifies streaming acoustic signals at up to – 3.5M data elements/sec

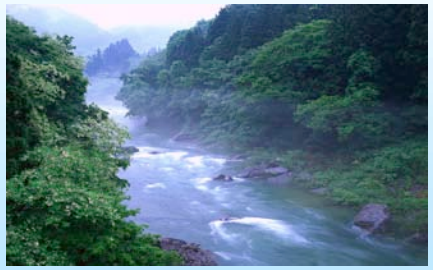
Cyber-security system analyzes massive amounts of network traffic to identify anomalies

Sensors in buoys provide faster and more accurate flood prediction

Real-time data streams help predict traffic patterns

Imagery is analysed in real-time to detect events, persons, or items of interest

Real-time environmental data detects changes to water resources



Big Data – Better Context

Sense Making: We understand something better by taking into account the things around it…

"@Steve Rocked The Slopes Today!"
1 minutes ago

[Hardly actionable]

"@Steve Rocked The Slopes Today!"
1 minutes ago

Work Comp Claim

Back Injury

Dr. Blacklist

[Substantially more actionable]

Context Accumulation: The incremental process of integrating new observations with previous observations.

Analytics that listen, measure and analyze social media performance to more effectively:

Enhance Service Outcomes

Understand citizens' needs to target new services cost-effectively through different social media channels

Enhance citizen relationships

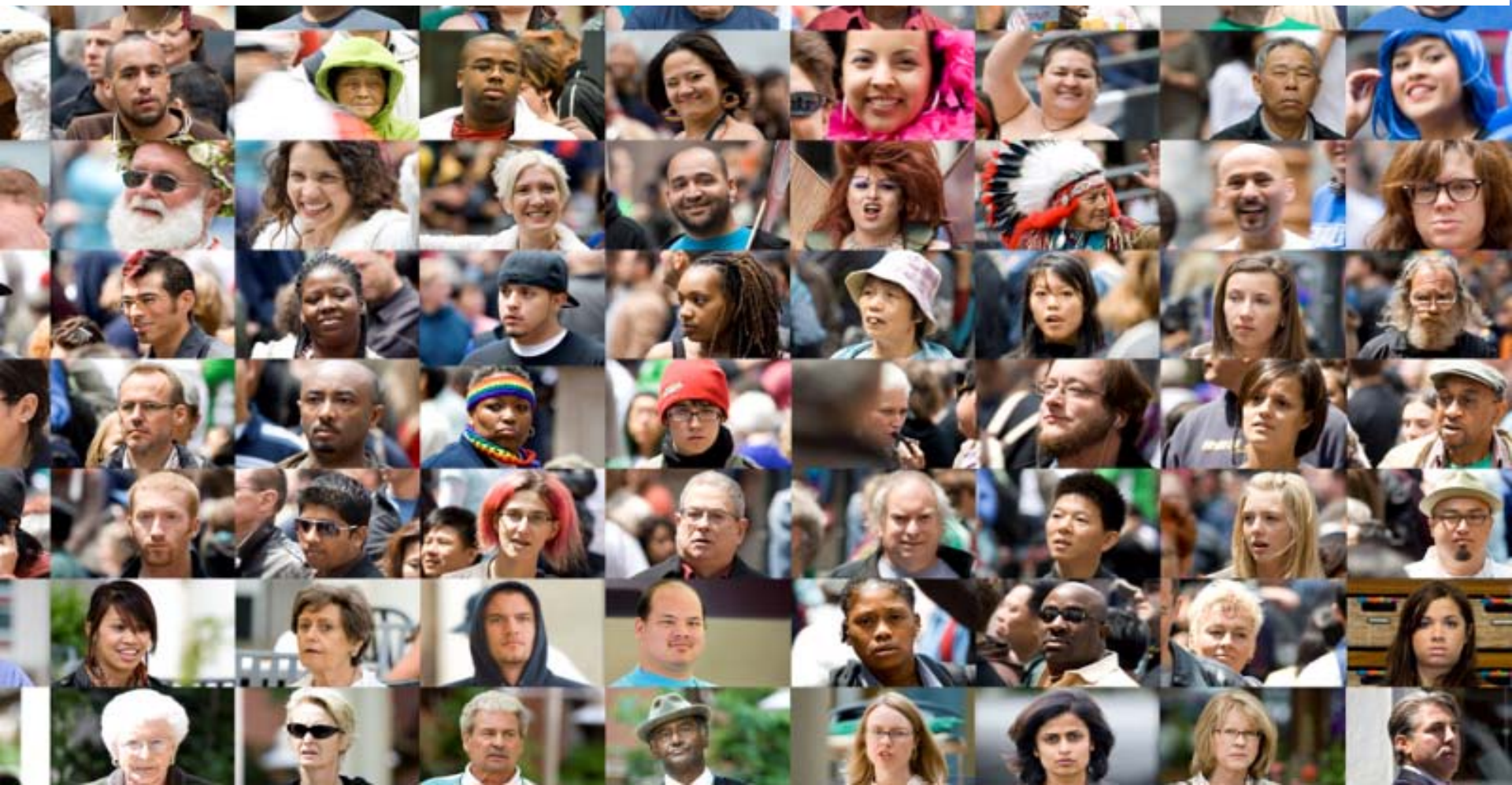
Evaluate your reputation and make evidence-based decisions that target the right stakeholders at the right time

Improve citizen experience

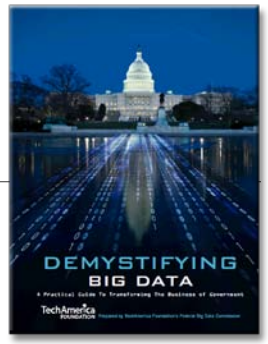
Respond more quickly with accurate, timely and relevant insight into citizens requests to ensure a consistent experience across all channels

Create Relationships. Build Advocacy. Improve Service.

Where/How Do We Start?

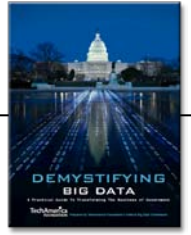


Tech America Big Data Report Findings



1. **Understand the "Art of the Possible"**
2. **Identify 2-4 key business or mission requirements that develop underpinning use cases that would create value for both the agency and the public.**
3. **Take inventory of your "data assets." Explore the data available both within the agency enterprise and across the government ecosystem within the context of use cases.**
4. **Assess your current capabilities and architecture against what is required to support your goals**
5. **Explore which data assets can be made open and available to the public to help spur innovation outside the agency.**

Practical Big Data Roadmap



Define

Define the Big Data opportunity including the key business and mission challenges, the initial use case or set of use cases, and the value Big Data can Deliver

Assess

Assess the organization's currently available data and technical capabilities, against the data and technical capabilities required to satisfy the defined set of business requirements and use cases

Plan

Select the most appropriate deployment pattern and entry point, design the "to be" technical architecture, and identify potential policy, privacy and security considerations

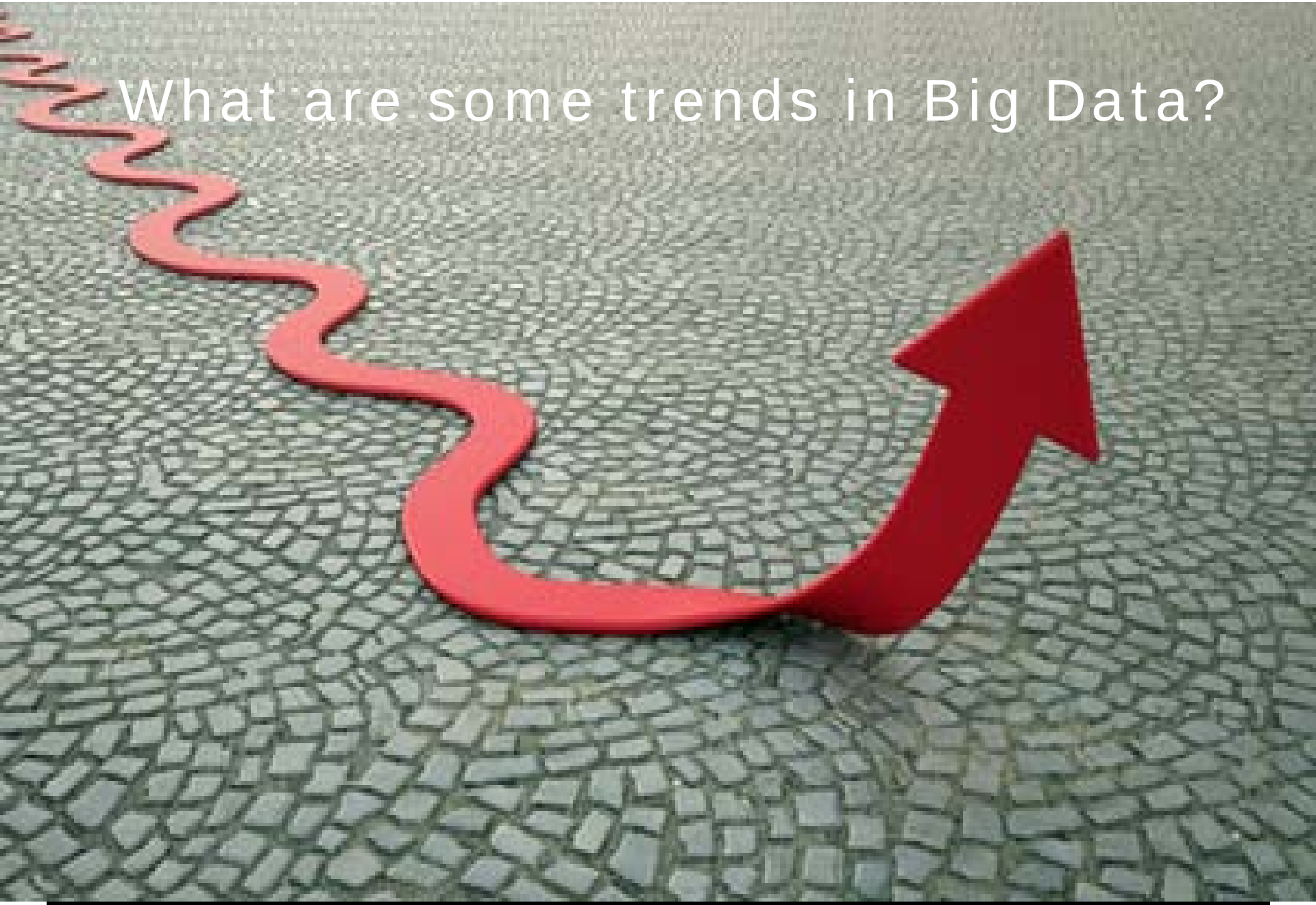
Execute

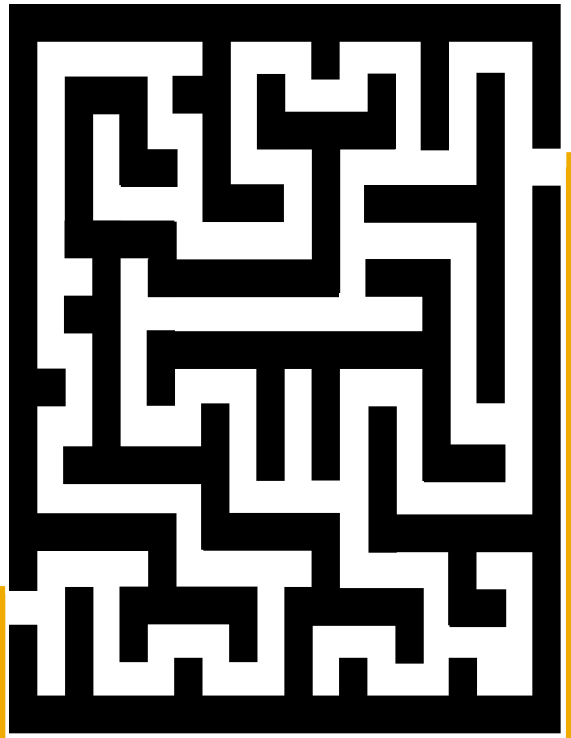
Deploy the current phase Big Data project, maintaining the flexibility to leverage its investment to accommodate subsequent business requirements and use cases

Review

Continually review progress, adjust the deployment plan as required,and test business process,policy, governance, privacy and security considerations

What are some trends in Big Data?





FASTER

SIMPLER

Open Data

Analytics for non analytics people

Mobile Analytics

Sentiment Analysis

Combined with… Cloud, Mobile

Predictive

Real-time

What are some Public Sector Relevant Use Case Examples?

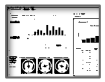
Changing citizens' lives through innovative organizations

MKI

Faster
Genome analysis

Better
Insight to support the needs of cancer patients in real-time

Greater
Personalization to individual patient needs



Leverages **R, Hadoop, in-memory**

MRI - Tokyo

Better traffic flow via on-demand modeling to redirect traffic via multiple applications

Incorporates **data from multiple sources** including real-time traffic, construction and road closures

Improved livability of city

Recovery.Gov

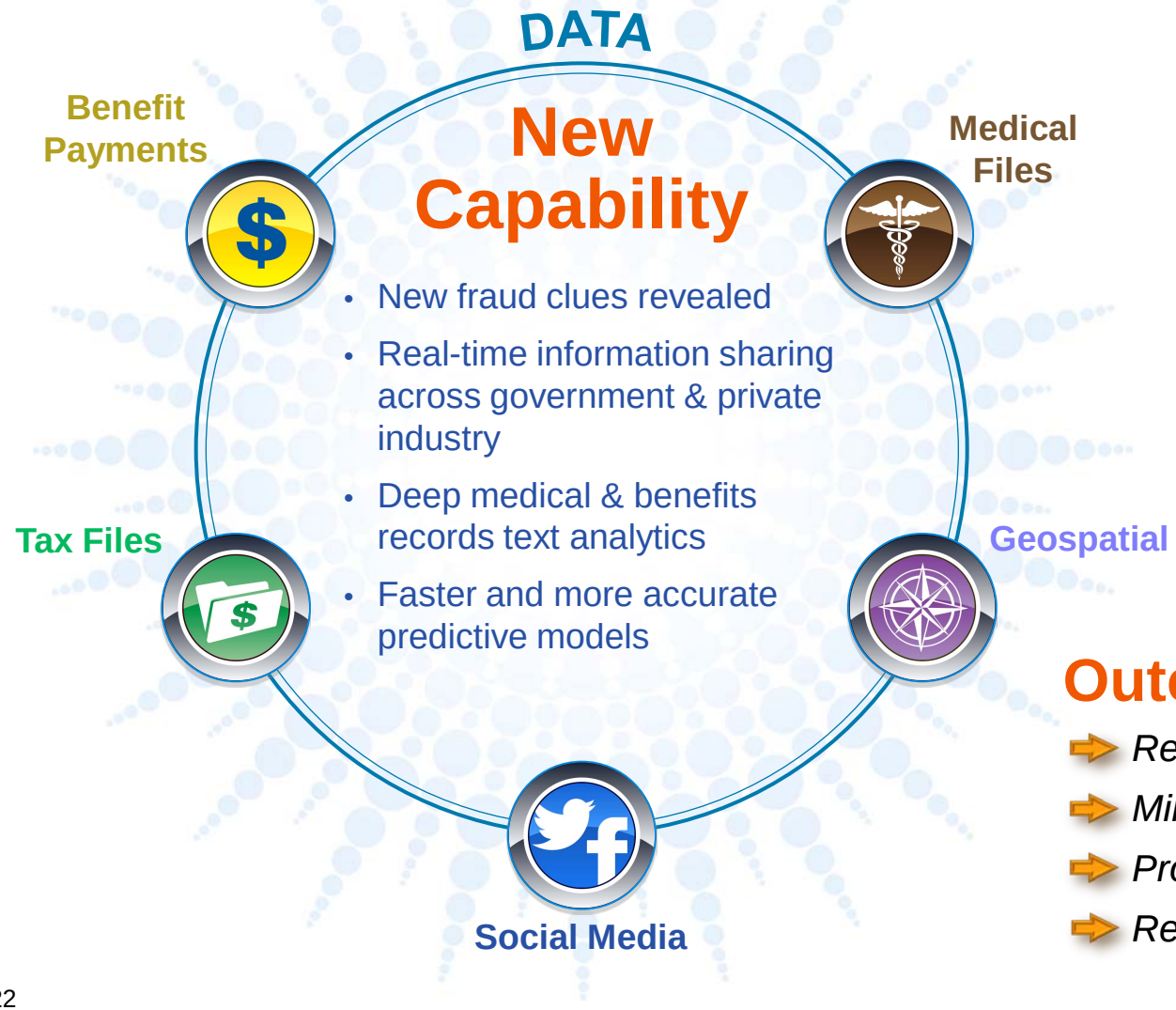
Transparency and accountability

Leverages **cloud, mobile, and analytics**

Analytics **for Non-Analytics People**

Tax and social program fraud, abuse and errors

An integrated approach to fighting fraud, abuse and error in tax and social programs



Outcomes

- *Reduce overpayments*
- *Minimize tax gap*
- *Proactively detect & deter fraud*
- *Reduce analysis time*

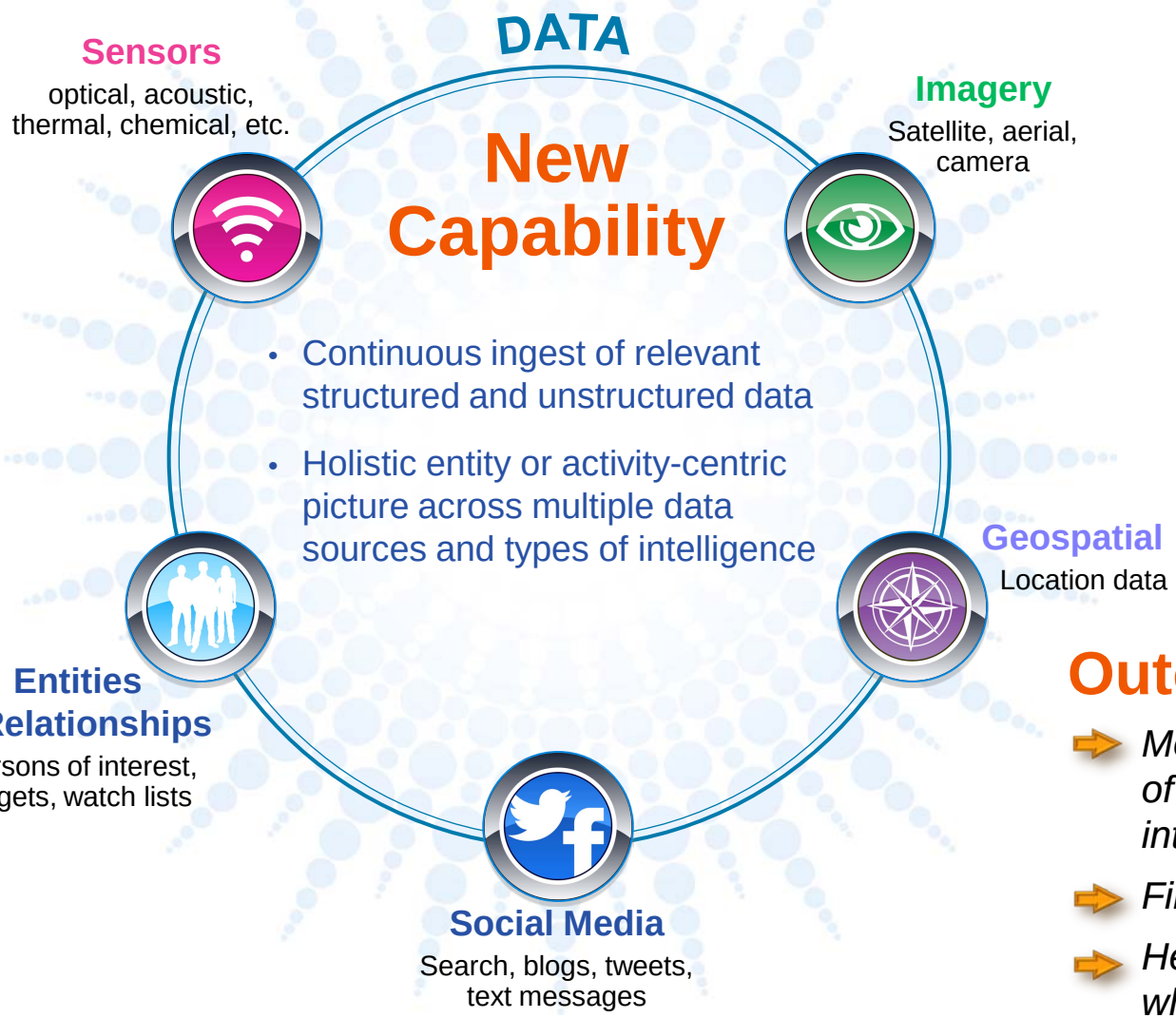
Reducing Fraud and Enabling Better Outcomes

Major government medical and social benefits agency

- Identified an improper payment level for a particular benefit of over 40%, *worth over $140 Million*
- Performed analysis in hours, instead of weeks
- Ad-hoc analysis of over 70 data sources, including: in-patient, out-patient, prescriptions, financial records, notices of death, criminal data, many others
- Utilizes analytic data warehouse appliance

Threat & Crime prediction and prevention

Identify and respond to threats and crime before it materializes

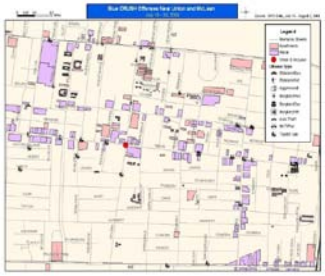


Outcomes

- *More reliable understanding of a suspect, target or area of interest*
- *Finds the dots, connects them*
- *Helps analysts understand what they don't know*

Threat and Crime Prediction & Prevention

Memphis Police Department



- Recognize crime trends *as they are happening*; enables changing tactics and redirecting resources before crime happens
- Integrates heterogeneous data, statistical modeling/analysis and GIS
- *30% reduction* in serious crime overall; *36% reduction* in one targeted area

TerraEchos

U.S. High Security Facility

- Needed a physical intrusion detector system able to detect, classify, locate and track potential threats – above and below ground
- Data arrives at the extremely *high data rate of 1.6 GB per second* and is processed and transmitted in real-time
- Sensitive enough to distinguish between a animal and an intruder
- Uses stream computing platform

I suggest we each put our Big Data link on this page

ore information:

ibm.com/bigdata