Big data and corrections: what’s the big issue?

Corrections Technology Association
June 4, 2013
Big is a relative term

What is considered "big data" varies depending on the capabilities of the organization managing the set.

- For some organizations, facing hundreds of gigabytes of data for the first time may trigger a need to reconsider data management options.

- For others, it may take tens or hundreds of terabytes before data size becomes a significant consideration.
Corrections by the numbers

1 in 33 American adults is under correctional control

PEW Center on the States – June 2012, The High Cost of Corrections in America: Infographic
 Corrections by the numbers

PEW Center on the States – June 2012, The High Cost of Corrections in America: Infographic
Corrections by the numbers

PEW Center on the States – June 2012, *The High Cost of Corrections in America: Infographic*
At year’s end 2011, nearly 1 in every 34 adult residents in the U.S. was under some form of correctional supervision* (approximately 7 million)

Assume each individual has a circle of 5 contacts worth noting; approximately 35 million individuals to be aware of (relationship mapping)

More than 10% of the population

*Correctional Populations in the United States, 2011 (Lauren E. Glaze, BJS Statistician, and Erika Parks, BJS Intern)
Justice in the data

Applying a data-driven lens to improve the efficiency and effectiveness of the criminal justice system has a potentially huge incentive:

Improving the bottom lines of state budgets around the country
Making decisions

Prisoner population projections

Current forecast: 16,509

Historical population: 14,234

Options:
- Option 1: 16,380
- Option 2: 15,381
- Option 3: 14,005

Source: Oregon Department of Corrections

DAN AGUAYO/THE OREGONIAN
Corrections ranks among the highest expenditures in state budgets

In 2012, corrections was No. 4 in California State budget

It costs $45,000 per year in New Jersey to incarcerate someone

The costs of incarceration just prior to trial alone is $9 billion

Cost of local systems nationally is estimated by the Department of Justice at more than $130 billion a year
The four V’s: understanding the trend

**Volume**
Typically, big data refers to high volumes of information

**Velocity**
Like water from a fire hose, data is constantly flowing into agencies at intense rates

**Value**
High volumes of data produced at accelerated rates

**Variety**
Non-traditional data sets associated with big data, such as video and audio files, structured and unstructured content, spatial tweets, blogs, wiki, and social media entries
Big data in corrections: where’s the data?

Within prisons and jails: **structured** data
- Inmate cell assignments, assessments, money transfers and purchases, rules violations, associates, property, program enrollments/outcomes, grievances, health appointments and encounters, pill line….
- Incidents, consumables, asset management, payroll, system/Internet usage

Within prisons and jails: **unstructured** data
- Document imaging, email, SharePoint, PDF reports
- Photos: mugshots, tattoos, evidence, visitors, staff
- Video: surveillance, committee hearings, visits, court hearings, health services
- Audio: inmate telephone conversations, hearings
- Biometrics

Add the following for community corrections:
- GPS tracking, case notes, co-workers and associates…
- Internet usage, cellular records (parolee and staff), email, social media…
Corrections in the digital age

Social Media and the Internet
Cyber threats and cyber supervision
Social networking

Supervising offenders online
Securing facilities

Investigating online
Social media policy

Supervising staff online

justice.vic.gov.au/socialmedia
Statistics sourced from: ComScore State of the Internet, February 2011
Investing in a comprehensive information management strategy will result in:

better early warning, to enable faster response  
real-time awareness, to know what’s happening on the ground now  
real-time feedback, perhaps “most important,” to see what’s not happening versus what was intended
Information overload

It’s happening now!
The returns for better applying technology in criminal justice extend far beyond reducing crime or costs, to something that government officials are sworn to uphold: justice.

Anne Milgram
Former Attorney General of New Jersey
What is big data?
How much is too much?
A fundamental shift is underway

New technologies
- Tools for "meaning-based computing" and real-time analytics processing
- HW and Appliances to power, process and store big data
- Cloud and Security for emerging demands to share information
- System integration services to help clients align, architect and accelerate the shift

Core Transaction Systems
Other Operational Systems
Analytical Environment

Predefined reporting, dashboards and analytics to:
- Measure/monitor the business
- Analyze and improve operations

“Human-friendly Information”
- Mobile
- Audio
- Texts
- Social Media
- Video
- Images
- Email

Requiring filters for meaning (context, relevance, urgency) to:
- Operate efficiently
- Protect the population
- Improve best practices through trends and public policy

Big Data
- Velocity
- Volume
- Variety
- Value

New technologies
- Tools for "meaning-based computing" and real-time analytics processing
- HW and Appliances to power, process and store big data
- Cloud and Security for emerging demands to share information
- System integration services to help clients align, architect and accelerate the shift
Leverage data explosion

Harness the Power of Information

Empowering enterprises to leverage 100% of information to drive faster intelligent business decisions

Return on Information: Time to Value

100% of Information

Information Infrastructure
- CAPTURE
- STORE
- REPPLICATE
- SCALE

Information Management
- MANAGE
- SECURE
- GOVERN
- LEVERAGE

Information Insight
- SEARCH
- ANALYZE
- UNDERSTAND
- ACT
Techology innovation

… is one of the main drivers behind “Big Data Analytics”

<table>
<thead>
<tr>
<th>Technology</th>
<th>Feature</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale-out</td>
<td>Massive Parallel Processing, Shared-nothing Node fail-over</td>
<td>Linear scalability, critical SLA Industry standard components</td>
</tr>
<tr>
<td>Columnar</td>
<td>Structured data stored by column</td>
<td>Performance on large data volumes Effective Compression</td>
</tr>
<tr>
<td>In-memory</td>
<td>Data permanent in memory</td>
<td>Performance on data streams and cubes</td>
</tr>
<tr>
<td>In-database analytics</td>
<td>Execute close to data</td>
<td>Performance on large data volumes</td>
</tr>
<tr>
<td>Complex event processing</td>
<td>Analyse and react to stream event data</td>
<td>Near real-time response to business process</td>
</tr>
<tr>
<td>Search</td>
<td>Index and search unstructured and structured data</td>
<td>Analyze unstructured data</td>
</tr>
<tr>
<td>Human Language and Rich Media Processing</td>
<td>Text, audio, video analysis</td>
<td>Meaning based computing Sentiment rating, other ratings</td>
</tr>
<tr>
<td>Real-time Visualization</td>
<td>Graphic display of complex data</td>
<td>Human perception, faster results</td>
</tr>
<tr>
<td>Content management systems</td>
<td>Support variety of formats May not support SQL or ACID</td>
<td>Store and process unstructured data</td>
</tr>
</tbody>
</table>

Big Data Reference Architecture
Emerging technology addresses challenges

<table>
<thead>
<tr>
<th>Technology</th>
<th>Volume</th>
<th>Variety</th>
<th>Velocity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hadoop distributions</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Content management systems</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Real-time analytical RDBMS</td>
<td>Yes</td>
<td></td>
<td>Yes</td>
</tr>
</tbody>
</table>

**Storing data**
- NoSQL
- Hadoop distributions
- Content management systems

**Processing data**
- SQL
- Traditional RDBMS for OLTP/OLAP
- Real-time analytical RDBMS
The challenges of unmanaged information

“Missed opportunity”
- Social Media
- Video
- Email
- Texts
- Messages
- Audio
- Images
- Word, Excel

90%

“Increased risk”

“Cost and complexity”

Meaningful intelligent information insight requires analysis of 100% of information

10%
Imagine a world where a conversation with your information drives intelligent business decisions

Meaningful intelligent information insight requires analysis of 100% of information

Intelligent information insight across structured, semi-structured and unstructured data

Right time, intelligent decision making

Positive ROI (Return on Information)
Why is processing human information different?

Human Information is made up of ideas, is diverse, and has context. Ideas don’t exactly match like data does; they have distance. Human Information is not static – it’s dynamic and lives everywhere. Legacy techniques have all fallen short.
Analytics: Optimization of Labor and Post Positions

Data:
- Incident data
- Time of day
- Location within prisons
- Officer post coverage

Result:
Adjust shifts and optimize post positions

Analyze for patterns and trends

Data:
- Incident data
- Time of day
- Location within prisons
- Officer post coverage

Result:
Adjust shifts and optimize post positions
Analytics: monitoring and surveillance

**Unstructured Data:**
- Arrest reports
- Pre-sentence investigations
- Court documents
- Criminal history
- Psychological evaluations
- Incident reports,
- Grievance and appeal documents
- Education assessments
- Correspondence
- Inmate “requests to staff”

**Relationship Analysis**

**Result:**
- Identify safety risks
- Informed decisions in inmate cell and work assignments
- Control of movements
- Concept-based, context aware
- Real-time self learning
- Document training

**Meaning-based correlation**
Unstructured and structured analysis

Logical architecture
Big data discovery

![Diagram of big data discovery process]

**Structured and unstructured data samples**
- Web content
- Telemetrics
- Semi-structured
- Intelligent networks

**Discovery environment**
- Client data ingestion
- Analysis and discovery
- Insights lead to more questions
- Patterns insights and ideas

**Business operations**
- Efficiency
- Cost
- Compliance
- Risk

**Phases**
- Phase 1
- Phase 2
- Phase 3
- Phase 4

New and revised operational and analytical processes to be implemented in the enterprise.
Enabling the journey to information optimization

Stage 1 - Information Aware
- Transitioning from managing to exploiting information
- Recognition of information as a strategic enterprise asset
- Early stage of IO journey
- Primary focus at executive insight

Stage 2 - Information Empowered
- Emerging business, technology partnership
- App-focused vs. info-focused

Stage 3 - Information Enabled
- Analytics-driven decisions
- Business partners w/ IT
- Fully incorporating structured and unstructured data

Stage 4 - Information Driven
- Analytics-driven decisions
- Business partners w/ IT
- Fully incorporating structured and unstructured data

Stage 5 - Information Transformed
- Winning as information leader
- Ability to value info as a new model
- Info as strategic lever throughout enterprise

Strategy and value realization

Information technology
Enabling the journey to information optimization

Stage 1
Information Aware

Stage 2
Information Empowered

Stage 3
Information enabled

Stage 4
Information driven

Stage 5
Information transformed

Information technology

Strategy and value realization

Business enablement
“Bright spots” in corrections analytics

<table>
<thead>
<tr>
<th>Category of Data Mining/Analytics</th>
<th>Example “Use Cases”</th>
</tr>
</thead>
</table>
| **Efficiency and Cost Avoidance** | • Position and post coverage  
 • Inmate transportation optimization  
 • Incident avoidance, inhibit policy violations  
 • Rehabilitation program effectiveness |
| **Public Health** | • Patient safety  
 • Illness outbreak control |
| **Safety and Risk Avoidance** | • Relationship analysis for investigations, surveillance |
Challenges of climbing the maturity model

Variety, velocity, volume, time to value

90% of digital content created by 2015 will be mixed data types¹

75% of currently deployed data warehouses will not scale sufficiently to meet new information velocity and complexity of demands by 2016²

48% Increase in the projected volume of digital content worldwide¹

86% of corporations cannot deliver the right information, at right time, to support enterprise outcomes all of the time³

¹Source: IDC Predictions 2012: Competing for 2020
²Source: Gartner - The State of Data Warehousing in 2012
³Source: Coleman Parkes Survey Nov 2012
Information optimization enables…

Delivering meaningful outcomes through insights and analytics across all data

• Understanding and leveraging data as a primary strategic asset
• Implementing an information governance model that covers data ownership, metadata, security, and visibility of data
• Discovering patterns and developing insights through predictive analytics to help close the gap between strategy and execution
• Enabling better decisions and outcomes by providing the right information, at the right time, by the right method
## Speaker contact information

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Company</th>
<th>Phone</th>
<th>Email</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stephen McHugh</td>
<td>HP</td>
<td>908.420.7343</td>
<td><a href="mailto:stephen.mchugh@hp.com">stephen.mchugh@hp.com</a></td>
</tr>
<tr>
<td>Chris Litton</td>
<td>Sierra Systems</td>
<td>250.882.0207</td>
<td><a href="mailto:ChristopherLitton@SierraSystems.com">ChristopherLitton@SierraSystems.com</a></td>
</tr>
<tr>
<td>Diana Zavala</td>
<td>HP</td>
<td>703.728.6441</td>
<td><a href="mailto:diana.zavala@hp.com">diana.zavala@hp.com</a></td>
</tr>
<tr>
<td>John Ward</td>
<td>HP</td>
<td>916.801.9687</td>
<td><a href="mailto:john.ward2@hp.com">john.ward2@hp.com</a></td>
</tr>
</tbody>
</table>