Master Data Management & Big Data Analytics

Potential WaTech Services
Master Data Management (MDM)
Master Data: Gartner definition

“…the consistent and uniform set of identifiers and extended attributes that describes the core entities of the enterprise including

- customers,
- prospects,
- citizens,
- suppliers,
- sites,
- hierarchies, and
- chart of accounts.”
Historical Approaches

- Master Files
- Reference Data
- Look-up Tables
- Point-to-Point Interfaces
Master Data Management (MDM): Gartner definition

“… a technology-enabled discipline in which business and IT work together to ensure the

- uniformity,
- accuracy,
- stewardship,
- semantic consistency and
- accountability

of the enterprise’s official shared master data assets.”
MDM Technology Approaches

**Do It Yourself (DIY)**
- Purely code and database design
- Hybrid, e.g. with Microsoft MDM

**Domain Specific MDM products**
- Customer Data Integration (CDI)
- Product Information Management (PIM)

**Multi-Domain Master Data Management products**
- Market direction
- Efficiencies of scale, skills
- Flexibility to use data across subject areas
MDM at WaTech

• Partnered with OFM on P20W Project, starting 2012
  • Design/Build Longitudinal Database (preschool through work)
  • RFP for software, including robust Identity Matching

• Enterprise Informatica toolset
  • Multi-Domain Master Data Management (MDM)
  • PowerCenter, Advanced Edition (ETL, data integration, data governance)
    • Metadata Manager, Business Glossary
  • Data Quality
  • Data Integration Hub (Real-time data integration)
Architectural Styles

- **Consolidation**
  - Matches and physically stores a consolidated view of master data
  - Updated after the event and not guaranteed up-to-date. Authoring remains distributed
  - No publish and subscribe. Not used for transactions, but could be used for reference
  - For Reporting, Analysis and Central Reference

- **Registry**
  - Matches and links to create a “skeleton” system of record
  - Physically stores the global ID, links to data in source systems and transformations
  - Virtual consolidated view is assembled dynamically and is often read-only. Authoring remains distributed
  - Mainly for Real-Time Central Reference

- **Coexistence**
  - Matches and physically stores consolidated view of master data
  - Updated after the event and not guaranteed up-to-date. Authoring remains distributed
  - Publishes the consolidated view. Not usually used for transactions, but could be used for reference
  - For Harmonization Across Databases and for Central Reference

- **Transaction**
  - Matches and physically stores the up-to-date consolidated view of master data
  - Supports transactional applications directly — both new and legacy — typically through service-oriented architecture interfaces
  - Central authoring of master data
  - Acts as System of Record to Support Transactional Activity

---

[Diagram of Architectural Styles]

---

**WaTech**
Washington Technology Solutions
Deployment Styles

- **Golden Record**
  - Latest, best version of each attribute.

- **Federated, Contextual Record**
  - Conflicting values maintained

- **Third Party Data Matching Service**
  - Send identifiers to service (e.g. WaTech)
  - Get back new identifiers to contact trading partner
Should WaTech add MDM to Service Catalog?

Potential assets to leverage:
- Software investment
- Staff training
- Production experience

Services and approaches to consider:
- Golden Record within agency
- Client Hub for 2 or more agencies
- 3rd Party data matching service between agencies
Big Data
90% of today’s data has been created in just the last 2 years

Every day we create 2.5 quintillion bytes of data

Every 60 seconds there are:
- 72 hours of footage uploaded to YouTube
- 216,000 Instagram posts
- 204,000,000 emails sent

50,000 GB/second is the estimated rate of global Internet traffic by 2018

Volume
Scale of data

(Enough to fill 10 million Blu-ray discs)

1 in 3 business leaders don’t trust the information they use to make decisions

Veracity
Certainty of data

$3.1 trillion is the estimated amount of money that poor data quality costs the US economy per year

Velocity
Speed of data

80% of data growth is video, images and documents

Variety
Diversity of data

90% of generated data is “unstructured”

This includes tweets, photos, customer purchase histories and customer service calls
# Industry Adoption

<table>
<thead>
<tr>
<th>Status</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Already deployed</td>
<td>37%</td>
</tr>
<tr>
<td>In process of implementing</td>
<td>22%</td>
</tr>
<tr>
<td>Plans to deploy in next 12 months</td>
<td>18%</td>
</tr>
<tr>
<td>Plans to deploy in next 13-24 months</td>
<td>12%</td>
</tr>
<tr>
<td>Considering deploying</td>
<td>12%</td>
</tr>
<tr>
<td>Likely to implement in future</td>
<td>7%</td>
</tr>
<tr>
<td>No plans to deploy</td>
<td>20%</td>
</tr>
</tbody>
</table>

- **We have already deployed**
- **We are in the process of implementing**
- **We have plans to deploy**
- **We are considering deploying**
- **We are likely to implement**
- **We have no plans to deploy**

**Data projects vs. SMBs:**
- **80%** of enterprise organizations have already deployed or plan to deploy big data projects.
- **63%** of SMBs.

---

**Q.** Is your company currently implementing, planning or considering data-driven projects (i.e. devising strategies and projects to generate more value from existing data)?
## Factors Driving Interest in Big Data Analysis
What data sources or challenges are driving, or would drive, your organization's interest in doing big data analysis?

<table>
<thead>
<tr>
<th>Factor</th>
<th>2015</th>
<th>2014</th>
</tr>
</thead>
<tbody>
<tr>
<td>Finding correlations across multiple, disparate data sources</td>
<td>48%</td>
<td>43%</td>
</tr>
<tr>
<td>Analyzing high-scale machine data from sensors, web logs, etc.</td>
<td>28%</td>
<td>23%</td>
</tr>
<tr>
<td>Predicting customer behavior</td>
<td>46%</td>
<td>44%</td>
</tr>
<tr>
<td>Identifying computer security risks</td>
<td>29%</td>
<td>25%</td>
</tr>
<tr>
<td>Predicting product or service sales</td>
<td>40%</td>
<td>36%</td>
</tr>
<tr>
<td>Analyzing web clickstreams</td>
<td>24%</td>
<td>26%</td>
</tr>
<tr>
<td>Predicting fraud or financial risk</td>
<td>32%</td>
<td>27%</td>
</tr>
<tr>
<td>Other</td>
<td>1%</td>
<td>3%</td>
</tr>
<tr>
<td>Analyzing social network comments for consumer sentiment</td>
<td>29%</td>
<td>24%</td>
</tr>
<tr>
<td>Big data analytics is not of interest to my organization</td>
<td>12%</td>
<td>14%</td>
</tr>
</tbody>
</table>

---

**WaTech**
Washington Technology Solutions

Big Data Landscape
Industry Staffing for Big Data Initiatives

- **Database Programmer**: 58%
- **Business Analyst**: 53%
- **Data Analyst**: 50%
- **Engineer**: 44%
- **Data Architect**: 43%
- **Research Analyst**: 31%
- **Director or Manager of Analytics**: 26%
- **R&D Specialist**: 26%
- **BI Director**: 26%
- **Data Visualizer**: 23%
- **Chief Data Officer**: 18%
- **Data Scientist**: 18%
- **Chief Analytics Officer**: 16%
- **Chief Digital Officer**: 13%

**CURRENT SKILL SETS**

**PLANS TO HIRE FOR THIS SKILL SET IN THE NEXT 12-18 MONTHS**

Sources:
- **BDG Enterprise Big Data Study, 2014**
- **Results**
- **Conclusion**
- **Demographics**
Big Data & Analytics Approaches

Do It Yourself (DIY)
- Buy or build tools, design, train, implement
- On-prem, in-cloud or hybrid

Partner with WaTech or Other Organization
- Leverage tools, expertise
- On-prem, in-cloud or hybrid

Outsource
- Contract to meet needs
- On-prem, in-cloud or hybrid
Data Management & Analytics at WaTech

- **Data Management**
  - 3,000+ Online Databases (OLTP)
  - Enterprise Relational Data Warehouses (e.g. AFRS)
  - Enterprise Dimensional Data Warehouses (e.g. HRMS)

- **Business Intelligence, Reporting and Analytics**
  - Self-Service Business Intelligence (SAP, Microsoft)
  - Platform as a Service: (SAP Business Objects)
  - SAS Visual Analytics
Should WaTech add Big Data & Analytics to Service Catalog?

Potential to leverage considerable:
• Software investment
• Staff training
• Production experience

Services and approaches to consider:
• Analytics Platform (e.g. SAS Visual Analytics)
• Microsoft PowerBI (extension to Office 365)
MDM and Big Data Initiatives in Washington State Government

Master Data Management
- Potential mastering of Chart of Accounts
- Project at Department of Early Learning
- AIM Project at Health Care Authority

Big Data & Analytics
- PRISM tool at DSHS Research and Data Analytics
- AIM Project at Health Care Authority
Questions & Feedback

Contact: Doug Buster
Data & Business Intelligence Manager