Investigative Data Warehouse Program Overview

Presented by

May 4, 2006
Overview

• Access tools and data through standard desktop computer, FBI Intranet, & a browser session
• Interactive search services
  – Search based on free-form entered phrase
  – Search for subject terms such as name, DOB, SSN, telephone number
  – Search on metadata such as case number, document date, author, etc.
  – Capability to refine searches & set of search results
• Batch search services – search for subject terms
• More than 12,000 users – including Joint Terrorism Task Force (JTTF) members
• Collection of 53 government multi-source datasets including data from FBI, CIA, DOS, DHS, FinCEN
• Authority to Hold Data:
  – OGC and Privacy Impact Assessments (PIA) done for all data
  – Information Sharing Policy Board (ISPB) approves all data
  – Audit and Security Log Protocols are maintained across all data

May 4, 2006
Background

• Program started in FY02
• Program goals in FY03 OMB 300 Exhibit
  – Data from 40 sources
  – 5,000 users
  – 100 – 200 M records
  – 3-5 second response time for interactive searches
• Program achievements
  – Data from 53 sources
  – 12,432 user accounts
  – Data from ~ 500 M documents
  – 3-8 second median response time for interactive searches
  – Added Batch Search service
  – Special Project Team provided services to 5 task forces or operations
• Authority to Operate through 2008 for Integration/Test & Operational Environments
• Authority to Hold Data
  – OGC and Privacy Impact Assessments (PIA) done for all data
  – Information Sharing Policy Board (ISPB) approves all data

May 4, 2006
Growth in User Accounts

12,432 total users as of 4/12/06
All Field Offices & 23 Legal Attache Offices

IDW USER ACCOUNTS

May 4, 2006
# Users

<table>
<thead>
<tr>
<th>Breakdown of IDW Accounts</th>
<th>No. of Accts as of 02/22/2006</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>FBI Agent</td>
<td>3,858</td>
<td>32%</td>
</tr>
<tr>
<td>FBI Analyst/Support</td>
<td>3,044</td>
<td>25%</td>
</tr>
<tr>
<td>Federal Agency (Non-FBI)</td>
<td>1,541</td>
<td>13%</td>
</tr>
<tr>
<td>State Agency</td>
<td>313</td>
<td>3%</td>
</tr>
<tr>
<td>Local Agency</td>
<td>975</td>
<td>8%</td>
</tr>
<tr>
<td>Other LE &amp; Analyst users</td>
<td>2,248</td>
<td>19%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>11,979</strong></td>
<td><strong>100%</strong></td>
</tr>
</tbody>
</table>

May 4, 2006
**IDW Unique Users per Week**

- Between ~1,200 & ~1,600 unique users sign on in any given week

May 4, 2006
Interactive Transactions (Weekly)

- Between ~40,000 & ~60,000 interactive transactions in any given week
Batch Processing

- Between ~ 50 & ~ 150 batch jobs submitted in any given week
Number of IDW Batch Jobs

7,642 Batch jobs were submitted between 05/25/2004 and 3/17/2006 by a total of 891 different users for an average of 9 batch jobs per user.

May 4, 2006
Interactive Search Response Times (Weekly)

Average response time ~ 8 to 12 secs,
Median response time ~ 3-8 secs

RESPONSE TIMES

May 4, 2006
# Composition of Data

<table>
<thead>
<tr>
<th>Document Source</th>
<th>No. of Docs as of 03/17/2006</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>FBI Sources</td>
<td>164,271,091</td>
<td>28%</td>
</tr>
<tr>
<td>Non-FBI Sources</td>
<td>422,915,362</td>
<td>72%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>587,186,453</strong></td>
<td><strong>100%</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Document Type</th>
<th>No. of Docs as of 03/17/2006</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structured</td>
<td>542,939,167</td>
<td>92%</td>
</tr>
<tr>
<td>Un-Structured</td>
<td>44,247,286</td>
<td>8%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>587,186,453</strong></td>
<td><strong>100%</strong></td>
</tr>
</tbody>
</table>

May 4, 2006
Case References to IDW Services

• 23,759 references to IDW in ACS/ECF as of 3/17/2006
  – 87% referred to NSB related Case Classifications
  – References to IDW in FCI cases increasing
  – IDW contains only limited case data from Cyber, Civil Rights and Public Corruption
  – CID requests for IDW accounts has increased dramatically in last 2 months

• Sample of batch service job titles
  – Vetting for special events
  – Specific investigations
  – Threat assessments
Areas for Improvement

• Need capability for continuous operation given loss of primary site or capability at primary site
• Additional resources needed to meet for projected growth in number of users and volume of data
• Need more robust (responsiveness and capacity) backup and restore capability
• Need more robust test capability
Status

• Core capability completed
  – Interactive search
  – Batch search
  – Monitor & control capability to manage Quality-of-Service

• Improving service
  – Optimizing capacity & responsiveness given resource constraints
  – Transitioning to Extraction/Transformation/Load service

• Demand for service growing beyond expectations & base resources
  – Host more data (4x-8x current 4 TB)
  – Provide service to more users (~20K)
  – Provide continuous monitoring for updates
  – Provide geo-coded data
  – Integrate with emerging Enterprise services

May 4, 2006
Conclusion

• IDW has managed growth in users & data while maintaining Quality-of-Service
• IDW users are obtaining operational utility from the services provided by IDW
• Additional resources needed to meet greater than anticipated demand for services
• IDW services should be enhanced to provide data-related services
  – Provide services to reduce effort of integration & data migration
  – Provide functions needed to manage data quality

May 4, 2006