A Process for Data Requirements Analysis

David Loshin
Knowledge Integrity, Inc.
loshin@knowledge-integrity.com
October 30, 2007
Agenda

- The criticality of business information
- Data requirements analysis
- Conducting stakeholder interviews
- Key data discovery artifacts
Data Warehouse Information Perspective
Business Information

- Definition of Business Information
  - Data that has been collected and organized to support management of business operations and decision making.
  - Non transactional data e.g. “counts”
  - Must be actionable

<table>
<thead>
<tr>
<th>Aggregation</th>
<th>Users</th>
<th>Delivery</th>
</tr>
</thead>
<tbody>
<tr>
<td>Detailed Operational Data</td>
<td>Staff and line managers</td>
<td>Queries and detail listings</td>
</tr>
<tr>
<td>Aggregated Management Data</td>
<td>Mid-level and senior managers</td>
<td>Summary reports and scorecards</td>
</tr>
<tr>
<td>Summarized Internal &amp; External Data</td>
<td>Executive staff</td>
<td>Dashboards</td>
</tr>
<tr>
<td>Structured analytic data</td>
<td>Special purpose – marketing, business process analysis</td>
<td>Data mining, OLAP, Analytics, etc.</td>
</tr>
</tbody>
</table>
Applications Driven by Business Information

Customer Analysis
- Customer Profiling
- Targeted Marketing
- Personalization
- Customer Satisfaction
- Customer Lifetime Value
- Customer Loyalty

Vendor Analysis
- Supplier Management
- Vendor Management
- Shipping
- Inventory Control
- Distribution Analysis

Staff Productivity
- Call Center Utilization
- Call Center Optimization
- Production Effectiveness

Business Productivity
- Defect Analysis
- Capacity Planning
- Asset Management
- Resource Planning
- Financial Reporting
- Risk Management
- Enterprise Reporting

Behavior Analysis
- Trend Analysis
- Web Activity
- Fraud/Abuse Detection
- Participant Attrition
- Interaction Analysis
Types of Business Information

- Key Performance Indicators (KPI) – metrics for monitoring progress toward objectives
- Operational Performance Indicators (OPI) – metrics for monitoring business processes, operations and workflow
- Leading Indicators (LI) – metrics that contribute to rolled-up KPIs or OPIs
- Supporting Metrics (SM) – metrics supporting drill-through of KPIs, OPIs, and LIs
- Business Intelligence (BI) – analytics to understand customer behavior, patterns, trends and segmentation
Business Intelligence and Metrics

Measures for setting objectives for strategic initiatives

Key Performance Indicators

Measures for tactical activities

Operational Performance Indicators

Leading Indicators

Supporting Metrics

Supporting Data
What is Data Requirements Analysis?

- The Data Requirements Analysis Process is a standard set of procedures for identifying the data needs of a Data Warehouse system.
- The steps are analogous to traditional requirements analysis, but focused on data rather than functional needs.
Goals of Data Requirements Analysis

- Identify Data Requirements
- Establish Data Requirements Traceability
- Improve Data Quality
- Improve Information Quality
Goals of Data Requirements Analysis

- Strategic Requirements
- Information Requirements
- Technical Requirements

Requirements Prioritization

- Fact/Qualifier Matrix
- Source Systems
- Source/Target Mapping

Conceptual Data Model
Data Warehouse Information Perspective

- Iterative, top-down approach
  - Business decisions
    - Data needed to support those decisions
  - Interviews to identify information needs and metrics:
    - Performance Management – KPI
    - Operations Management – OPI
    - Business Analysis – BI
    - Customer Performance – CPI
  - Analysis
  - Specifications
1. Is the data knowable?

Knowable? →

No → Don't bother; just philosophy

Yes →

2. Is the data recorded on some media?

Recorded? →

No → In some human brain?

Yes →

3. Is the data accessible with reasonable cost and effort?

Access? →

No → Not worth the effort

Yes →

4. Is the data navigable? Can we find what we want?

Navigable? →

No → If you can't find it, you cannot use it.

Yes →

5. Is the data understandable? (language, technology, culture, etc.)

Understood? →

No → But can you find translation resources?

Yes →

6. Is the data of adequate quality?

Quality? →

No → Bad quality data can be misleading—worse than useless.

Yes →

7. Is the data relevant to our needs?

Relevant? →

No → Noise

Yes → Useful, valuable data.
## Driving Out Business Information Requirements

### Relevance

Does the data:
- Support business processes?
- Provide metrics for OPIs and KPIs?
- Answer business questions?
- Enable correct measures?
- Reflect real world activities or state?
- Provide useful capture points, timeliness

### Value-Added

Can the data improve:
- Operational visibility (e.g., document workflow and bottlenecks)?
- Efficiencies (e.g., in volume or processing times)?
- Decision making (e.g., expose patterns, trends, problems and opportunities)?

### Available

Does data support:
- Captured information requirements?
- Timeliness for updates to support information requirements?
- Organization structure to support information navigation needs?

---

**Data sources**

**Granularity**

**Dimensionality**
Data Discovery

Data Discovery tasks focus on:
- Capturing business information requirements, business processes, and terminology
- Identifying and defining the source data sets
- Defining the data quality dimensions for these data elements.

Deliverables
- Context Diagram
- Business Questions
- Fact/Qualifier Matrix
- Candidate Source Systems
- Source to Target Mapping
- DB Scenarios including Transformation Rules & Business Logic
The Requirements Analysis Process

- Identify the business context
- Conduct stakeholder interviews
- Synthesize requirements
- Source to Target mapping
Identifying the Business Context - Activities

1. Identify relevant stakeholders
2. Review and summarize the overall goals, objectives for the data warehouse
   - Acquire all project documentation
   - Acquire all system documentation
3. Review and summarize the scope of the reporting platform’s capabilities
4. Review and summarize impacts and constraints
Identifying Business Context - Diagram

ODS
Detailed level data
Reference tables

BI Repository
Metadata
Business Rules

Querying
Reporting
Analytics

Integration tables

Users review business information via query/reporting/analytics tools
Conduct Stakeholder Interviews - Activities

1. Identify interview candidates
2. Prepare and schedule interviews
3. Conduct interviews
4. Summarize interview notes
5. Identify information gaps for follow-up
## Summarize Interview Notes

<table>
<thead>
<tr>
<th>Interviewee:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Date/Time/Location:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Question:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Answer/Notes:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Question:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Answer/Notes:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>
Synthesize Requirements - Activities

1. Create a reference workflow model
2. Develop and validate the Fact/Qualifier Matrix
3. Identify candidate source systems
4. Identify and Standardize Common Business Terms
A “business process” is a coordinated set of activities intended to achieve a desired goal or produce a desired output product.

Models are designed to capture both the high level and detail of the business process.
Shared Data Objects

- Interactions between activities depend on shared data:

  - Instance values of common data types representing business facts communicate input and control during the business process.
  - Use the model to identify measurement points to be subjected to reporting and/or analysis.
### Work Flows and Tasks - Example

<table>
<thead>
<tr>
<th>Work Flow and Task</th>
<th>Event Trigger</th>
<th>Event Status</th>
<th>Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Work Flow A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Work Flow B</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task 2</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Facts and Qualifiers

- **Facts**: specific business questions or discrete items of information to be tracked/monitored/reported
  - Examples: counts, volumes, high water marks

- **Qualifiers**: conditions or dimensions used to filter or organize facts
  - Examples: regions, owners, time

<table>
<thead>
<tr>
<th>Facts</th>
<th>Qualifiers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Time</td>
</tr>
<tr>
<td>800# calls processed</td>
<td>X</td>
</tr>
<tr>
<td>Submissions received by mail</td>
<td>X</td>
</tr>
<tr>
<td>Submissions returned</td>
<td>X</td>
</tr>
<tr>
<td>Emails processed</td>
<td>X</td>
</tr>
</tbody>
</table>
# Common Business Terms - Glossary

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Address Verification</td>
<td>An employee verifies that the submitters address on file is correct for contact purposes.</td>
</tr>
<tr>
<td>Business Intelligence (BI) Repository</td>
<td>The BI Repository captures and maintains metadata (data about data). The metadata housed in the BI Repository consists of business and technical data names and definitions; information about where the described data resides in enterprise databases; and what database applications and reports use the data. The BI Repository is accessible through a web front-end application.</td>
</tr>
<tr>
<td>CMM</td>
<td>Capability Maturity Model describes the principles and practices, through which organizations define, implement, measure, control, and improve their software processes and is intended to assist organizations in software quality and process improvement.</td>
</tr>
<tr>
<td>Count (Workload)</td>
<td>For each Unit of Work, the counts associated at the task level for each data mart status (e.g., Receipt, Pending, Clearance, and Completion).</td>
</tr>
</tbody>
</table>
Proposing the Target Model

- Evaluate the catalog of identified data elements
  - Seek out the frequently created, referenced, modified, retired
- Assess object organizational structure
  - Evaluate conceptual structures as they map to business process use
    - Example: locations are composed of street, city, state, ZIP code
- Identify and resolve anomalies across data element sizes, types, formats
- Propose an object model
- Validate the object model within the information framework
- Validate the object model within the application framework
Source to Target Mapping - Activities

1. Identify and document source to target data mapping for each target element
2. Define the relevant transformation rules based on business scenarios
3. Identify the Candidate Reference Tables
# Source to Target Mapping – Example

<table>
<thead>
<tr>
<th>Target Data Element</th>
<th>Source Systems</th>
<th>Source Table</th>
<th>Source Data Elements</th>
<th>Transformations</th>
</tr>
</thead>
<tbody>
<tr>
<td>ProcessType</td>
<td>SYSA</td>
<td>SUBTABA</td>
<td>Process_Type</td>
<td></td>
</tr>
<tr>
<td>Order_Count</td>
<td>SYSA</td>
<td>SUBTABB</td>
<td>UID</td>
<td>( \text{COUNT(SYSA.UID)} )</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>\text{WHERE} SYSA.LNAME = SYSB.LASTNAME</td>
</tr>
<tr>
<td>SYSB</td>
<td>W3</td>
<td></td>
<td>LASTNAME</td>
<td></td>
</tr>
</tbody>
</table>
Data Assessment

- Tasks focus on:
  - Identifying the best sources of data
  - Assessing the quality of the data sources
  - Identifying gaps in requirements versus availability

- Deliverables:
  - Data assessment results
  - Updated Source to Target Map
  - Gap report
Data Assessment

Data Assessment Phase

- Source to Target Map
- Source System Metadata
- Quality Assessment Dimensions

Data Quality Audit → Data Assessment Results

Data Quality Gap Analysis → Gap Report

Best Candidate Selection

Updated Source to Target Map
Data Profiling for Data Requirements Analysis

- Characteristics of data element metadata is critical for analysis
- Combination of artifact review and empirical analysis
- Data profiling can provide “ground-truth” evidence of consistency with metadata
- Provides insight into suitability of candidate sources to satisfy the target needs
# Gap Analysis Template

<table>
<thead>
<tr>
<th>Target Data - Data Element Name</th>
<th>Source Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>System</td>
<td>Table</td>
</tr>
<tr>
<td>-------</td>
<td>------</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Summary

- **Inputs**
  - System documentation
  - Stakeholder interviews

- **Processing**
  - Capturing business information requirements, business processes, and terminology
  - Identifying and defining the source data sets

- **Outputs**
  - Context Diagram
  - Business Questions
  - Candidate Source Systems
  - Fact/Qualifier Matrix
  - Source to Target Mapping
  - Transformation Rules & Business Logic
  - Glossary
  - Business Process Diagrams
  - Business Process Descriptions
If you have questions, comments, or suggestions, please contact me

David Loshin
301-754-6350
loshin@knowledge-integrity.com