# 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

Data

Information

Knowledge

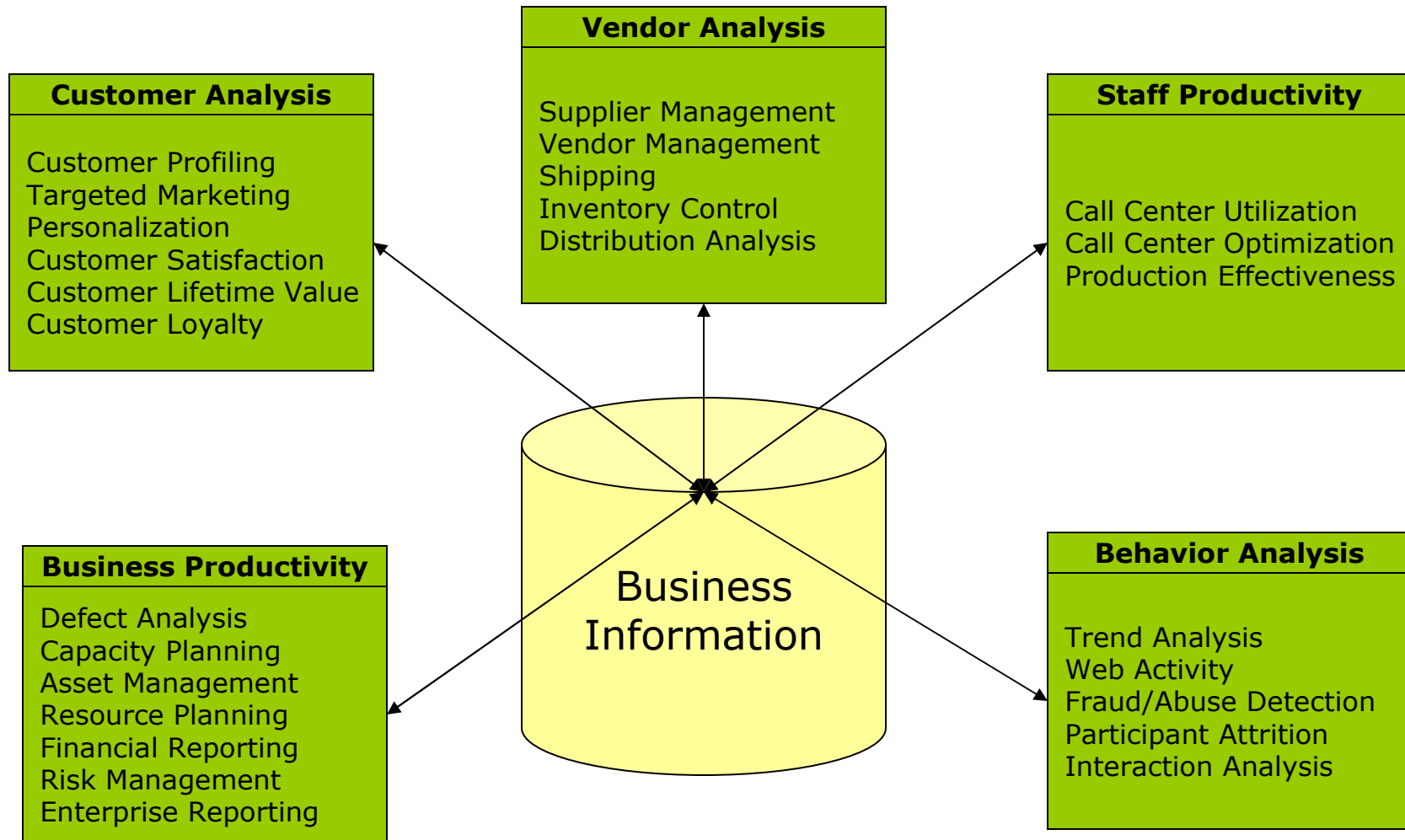Business Decision and/or Action
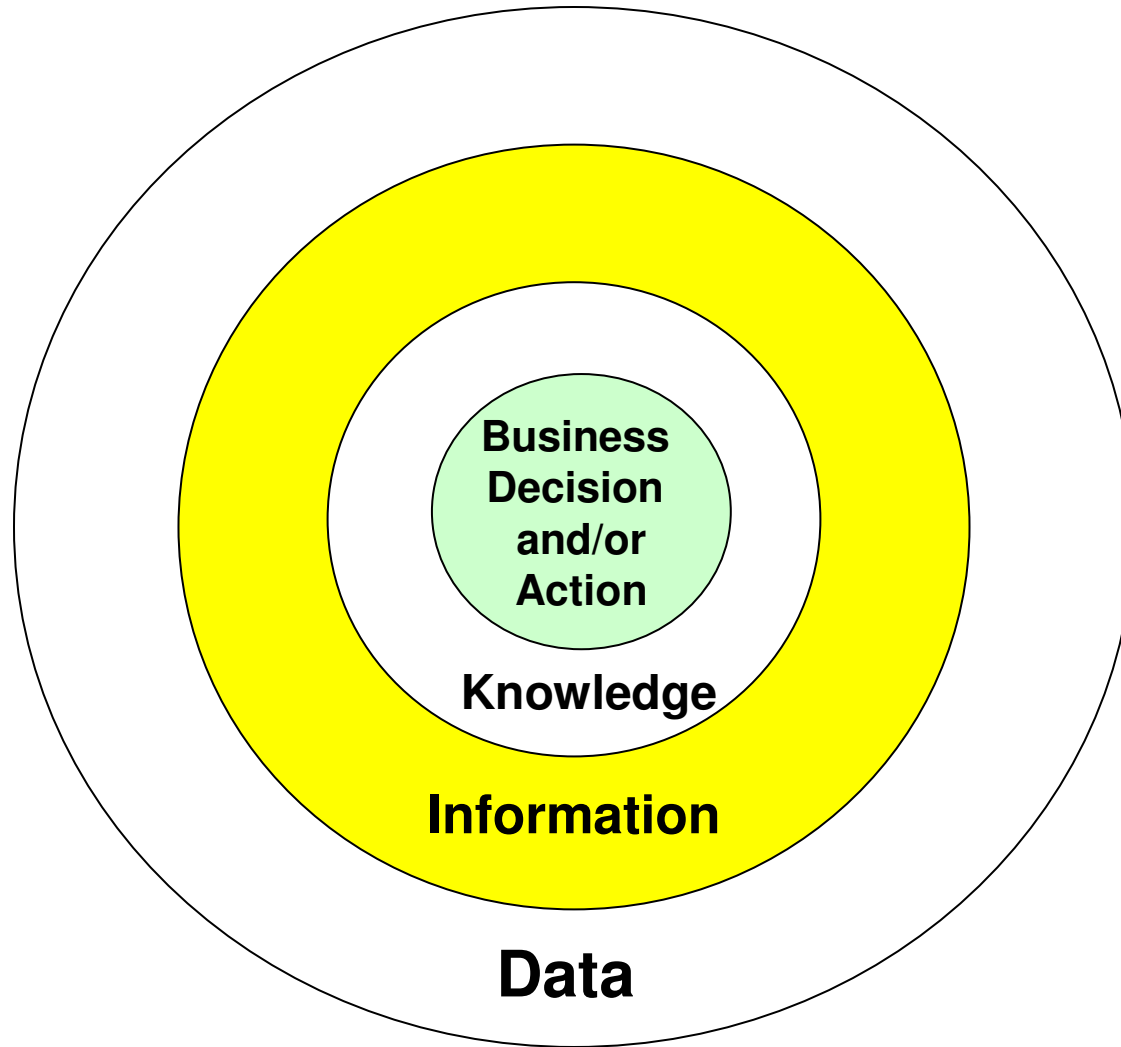
Business Information

- 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

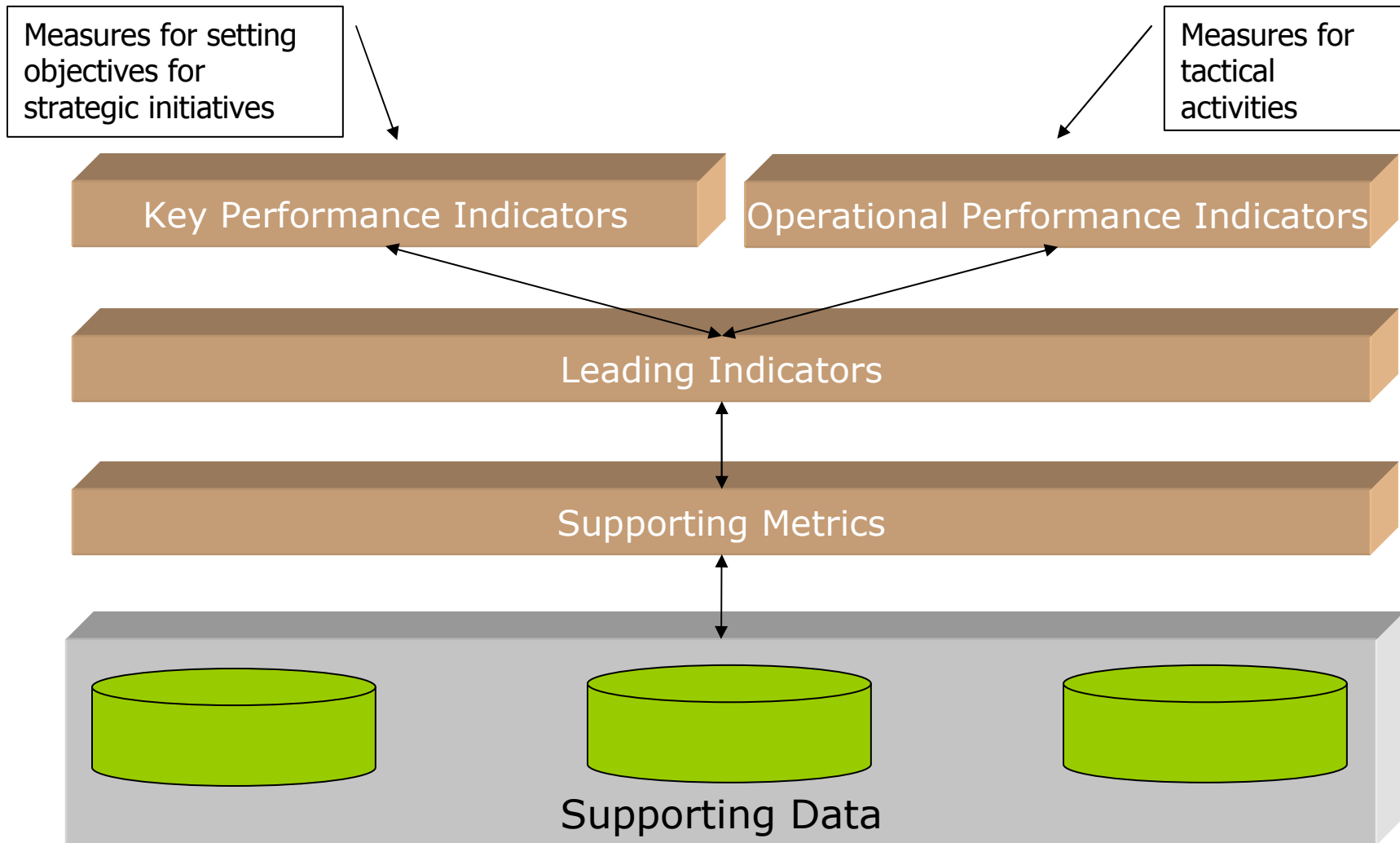| Aggregation	Users	Delivery
Detailed Operational Data	Staff and line managers	Queries and detail listings
Aggregated Management Data	Mid-level and senior managers	Summary reports and scorecards
Summarized Internal & External Data	Executive staff	Dashboards
Structured analytic data	Special purpose – marketing, business process analysis	Data mining, OLAP, Analytics, etc.

Applications Driven by Business Information

Vendor Analysis

Supplier Management
Vendor Management
Shipping
Inventory Control
Distribution Analysis

Customer Analysis

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

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
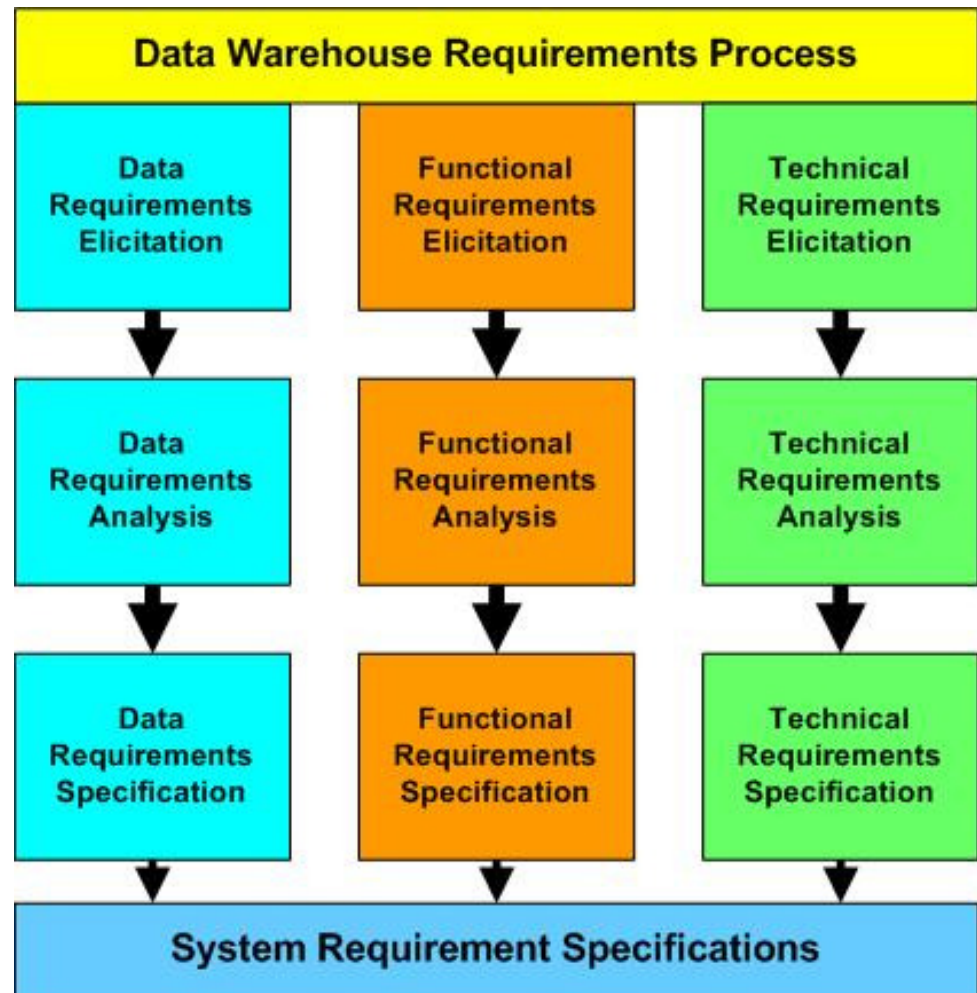
Business
Information

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

Measures for tactical activities

Key Performance Indicators

Operational Performance Indicators

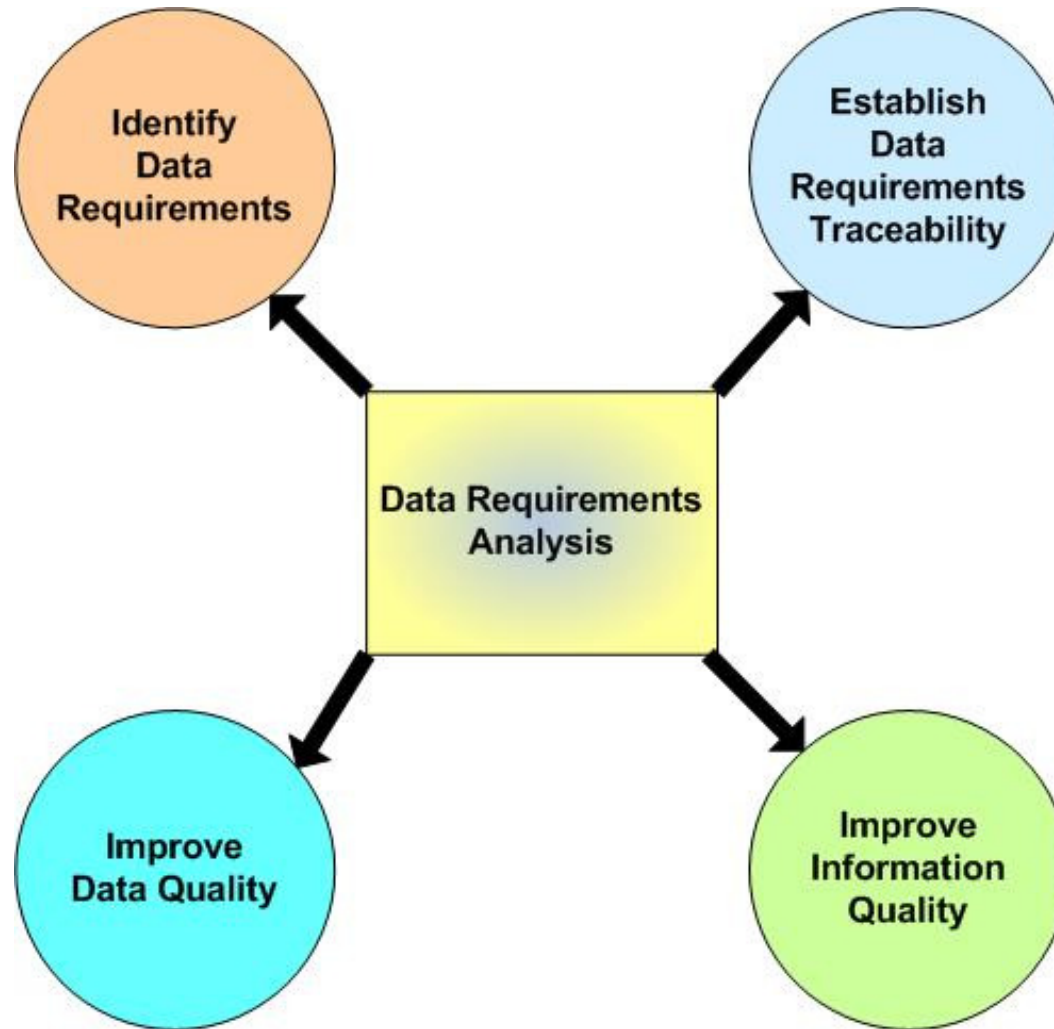Leading Indicators

Supporting Metrics

Supporting Data
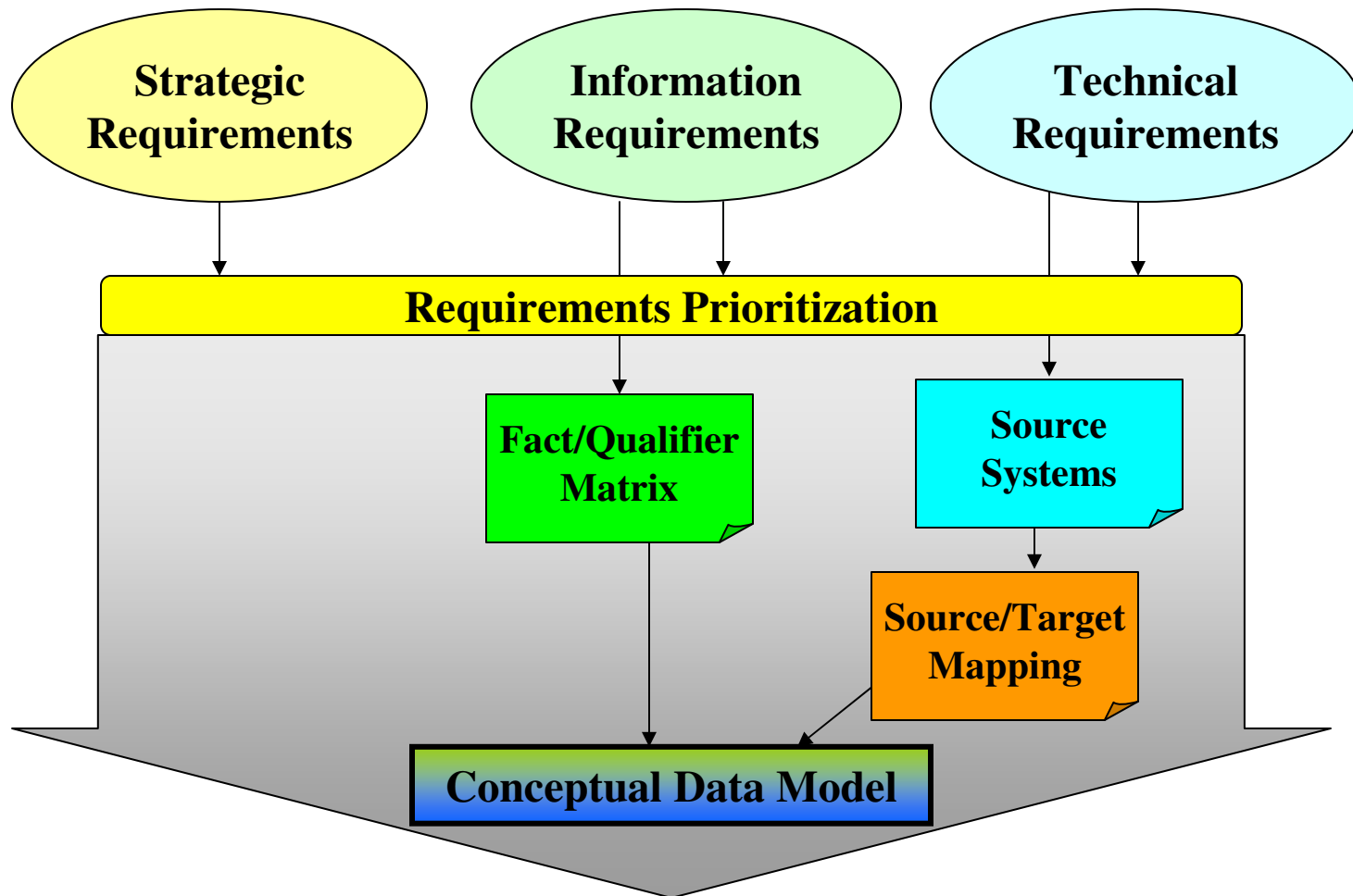
What is Data Requirements Analysis?

- 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.

Data Warehouse Requirements Process

Data Requirements Elicitation	Functional Requirements Elicitation	Technical Requirements Elicitation
Data Requirements Analysis	Functional Requirements Analysis	Technical Requirements Analysis
Data Requirements Specification	Functional Requirements Specification	Technical Requirements Specification

System Requirement Specifications
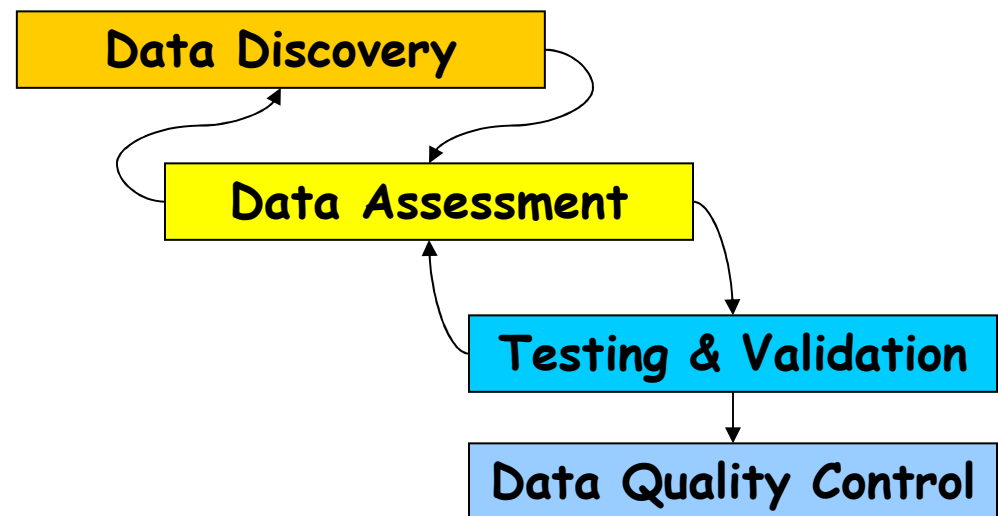
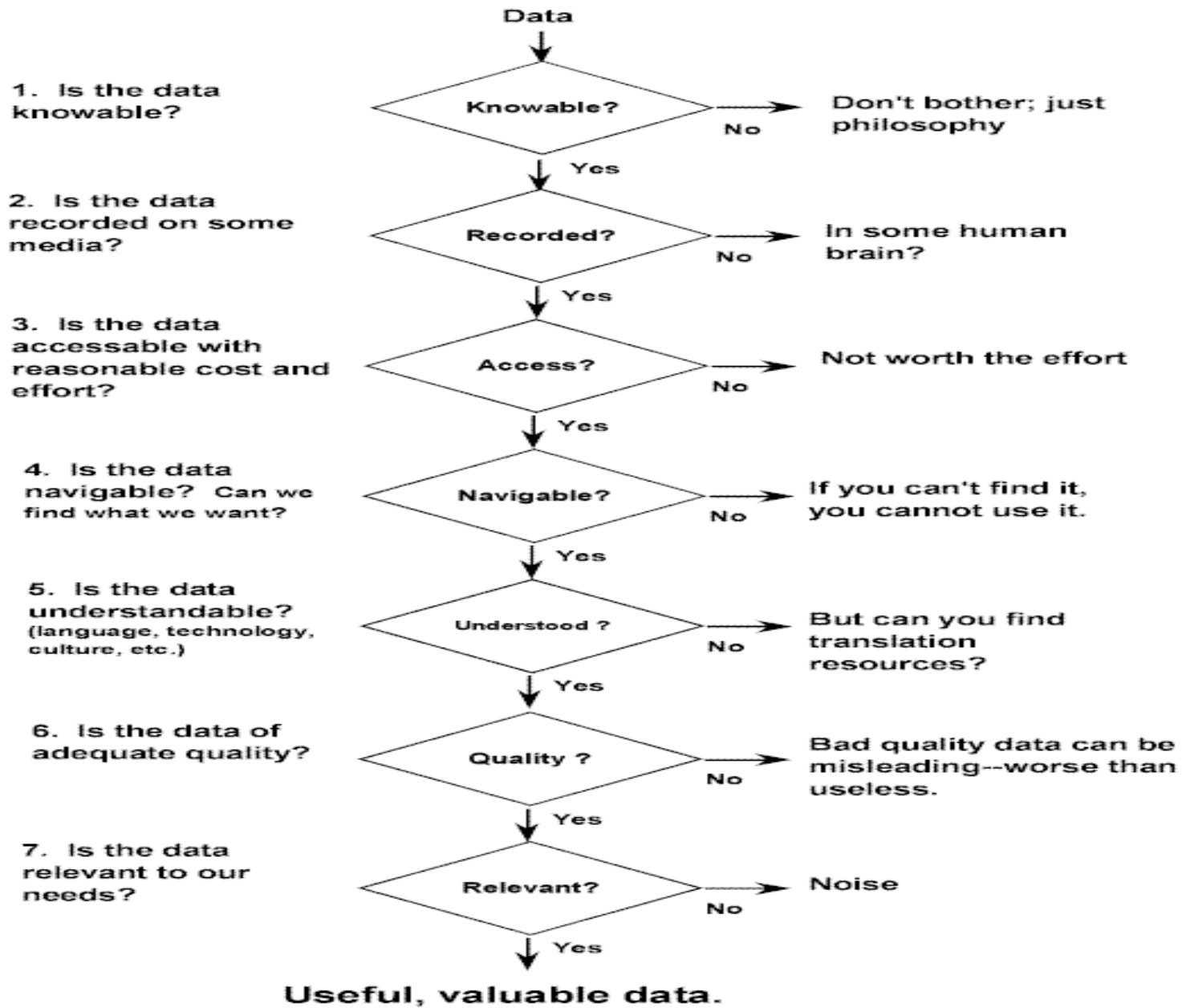# Goals of Data Requirements Analysis

Goals of Data Requirements Analysis

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

```
┌──────────────────────┐
│    Data Discovery    │
└──────────────────────┘
        ↕
┌──────────────────────┐
│   Data Assessment    │
└──────────────────────┘
        ↓
┌──────────────────────┐
│ Testing & Validation │
└──────────────────────┘
        ↓
┌──────────────────────┐
│ Data Quality Control │
└──────────────────────┘
```

Data

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 accessable 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.

12

Driving Out Business Information Requirements

| Relevance	Value-Added	Available
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	**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)?	**Does data support:** •Captured information requirements? •Timeliness for updates to support information requirements? •Organization structure to support information navigation needs?

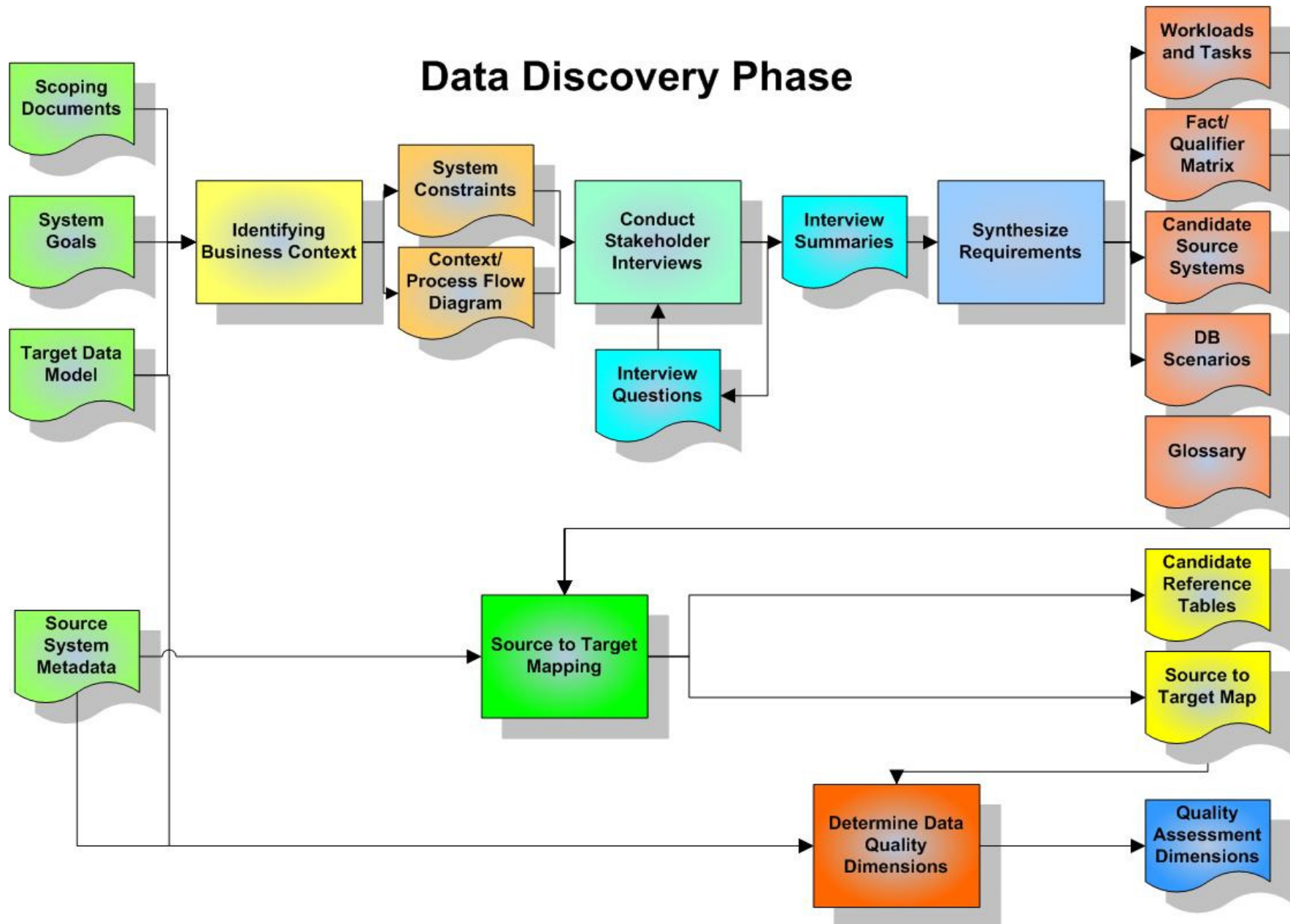Data sources
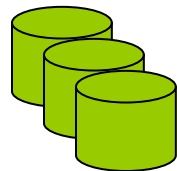
Granularity
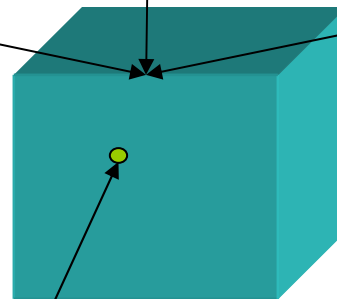
Dimensionality
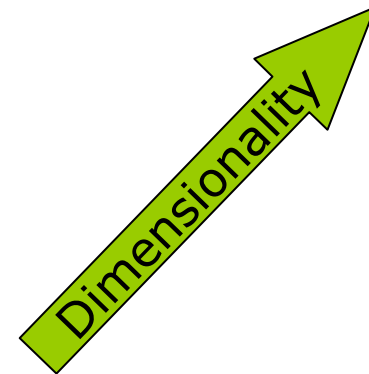
Data Discovery

- 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

Data Discovery Phase

Scoping Documents

System Goals

Target Data Model

Identifying Business Context

System Constraints

Context/ Process Flow Diagram

Conduct Stakeholder Interviews

Interview Questions

Interview Summaries

Synthesize Requirements

Workloads and Tasks

Fact/ Qualifier Matrix

Candidate Source Systems

DB Scenarios

Glossary

Source System Metadata

Source to Target Mapping

Candidate Reference Tables

Source to Target Map

Determine Data Quality Dimensions

Quality Assessment Dimensions

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
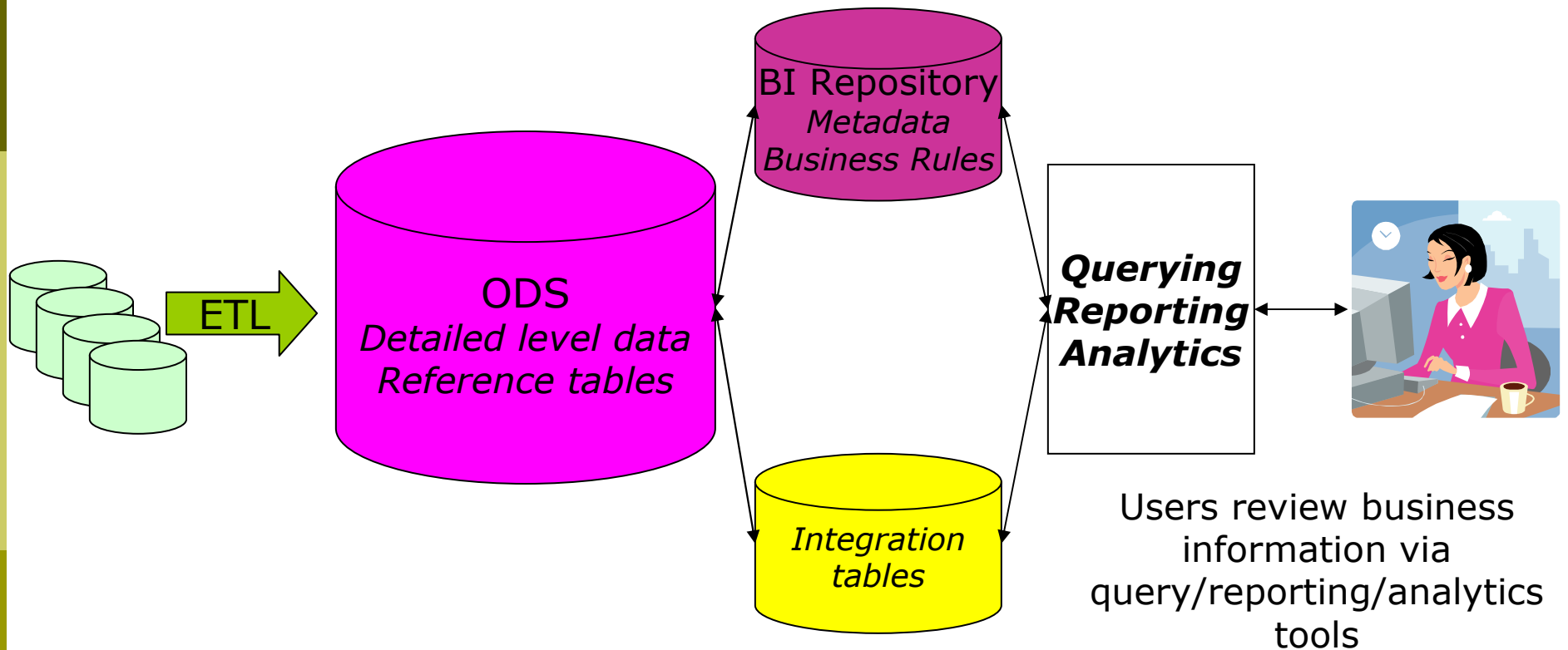
Conduct Stakeholder Interviews - Activities

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

Interviewee:	
Date/Time/Location:	

| |
|---|
| Question: |
| Answer/Notes: |
| Question: |
| Answer/Notes: |

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

Business Process Modeling

- 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
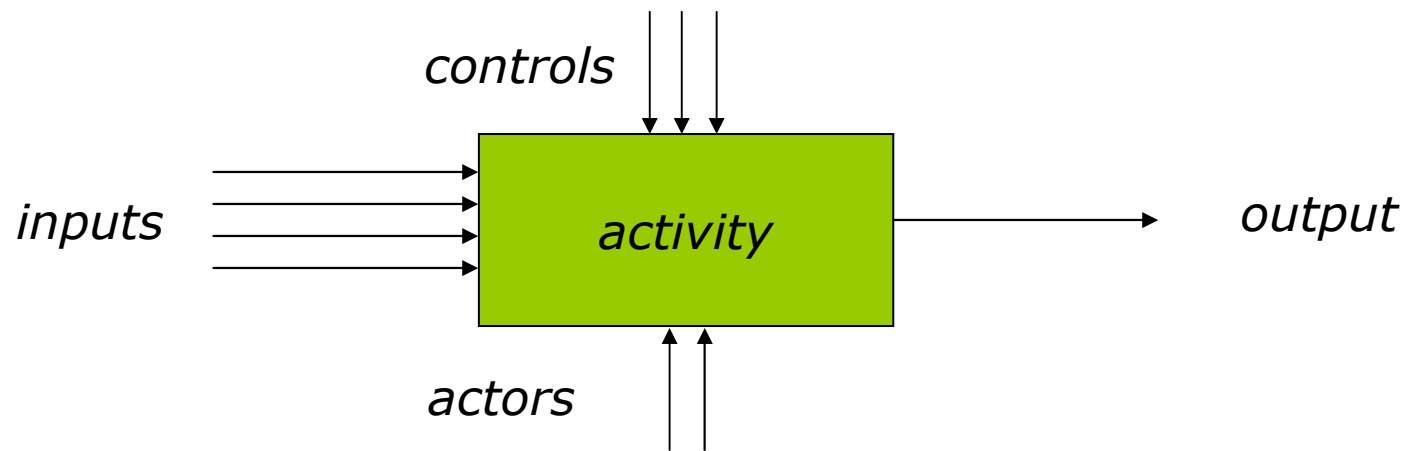
Shared Data Objects

- Interactions between activities depend on shared data:
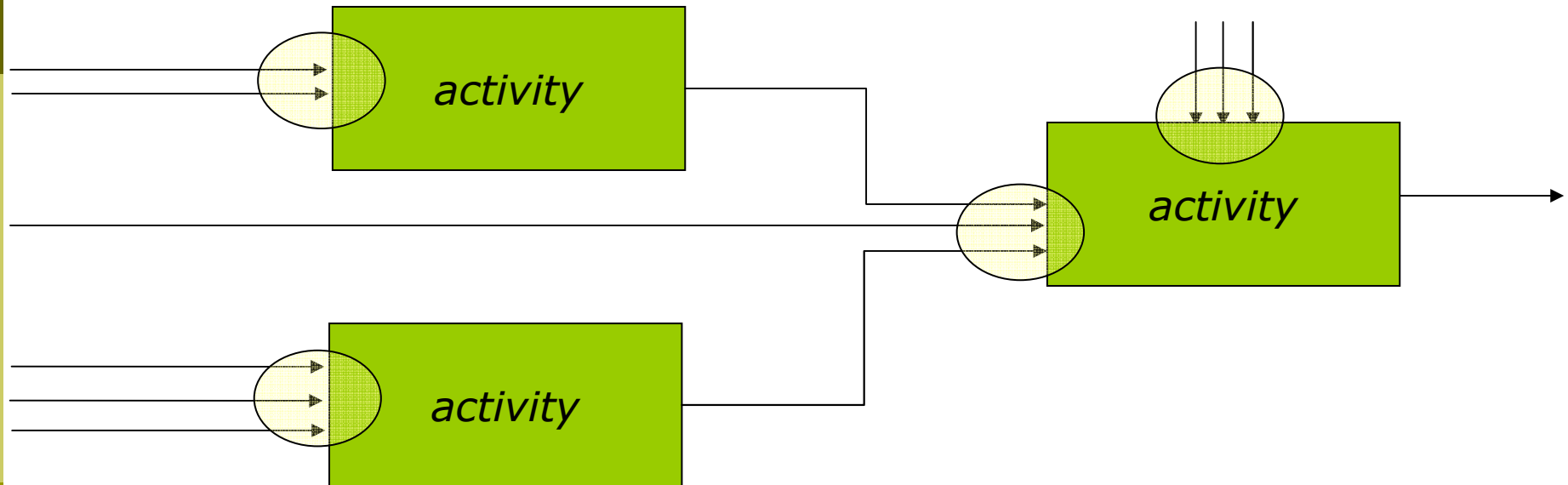
- Instance values of common data types representing business facts communicate input and control during the business process
- Use the model to identify measurement points to be subjected to reporting and/or analysis

Work Flows and Tasks - Example

Work Flow and Task	Event Trigger				Event Status				Measures			
	Initiated	Rejected	Rework	Complete	Received	Pending	Cleared	Complete	Volume	Start Date	End Date	Duration
Work Flow-A												
Task 1												
Task 2												
Work Flow-B												
Task 1												
Task 2												

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

Facts	Qualifiers			
	Time	Region	Media	Office
800# calls processed	X			X
Submissions received by mail	X	X	X	
Submissions returned	X	X	X	
Emails processed	X		X	X

Common Business Terms - Glossary

Term	Definition
Address Verification	An employee verifies that the submitters address on file is correct for contact purposes.
Business Intelligence (BI) Repository	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.
CMM	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.
Count (Workload)	For each Unit of Work, the counts associated at the task level for each data mart status (e.g., Receipt, Pending, Clearance, and Completion).

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

Target Data Element	Source Systems	Source Table	Source Data Elements	Transformations
ProcessType	SYSA	SUBTABA	Process_Type	
Order_Count	SYSA	SUBTABB	UID	*COUNT(SYSA.UID) WHERE SYSA.LNAME = SYSB.LASTNAME*
			LNAME	
	SYSB	W3	LASTNAME	

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
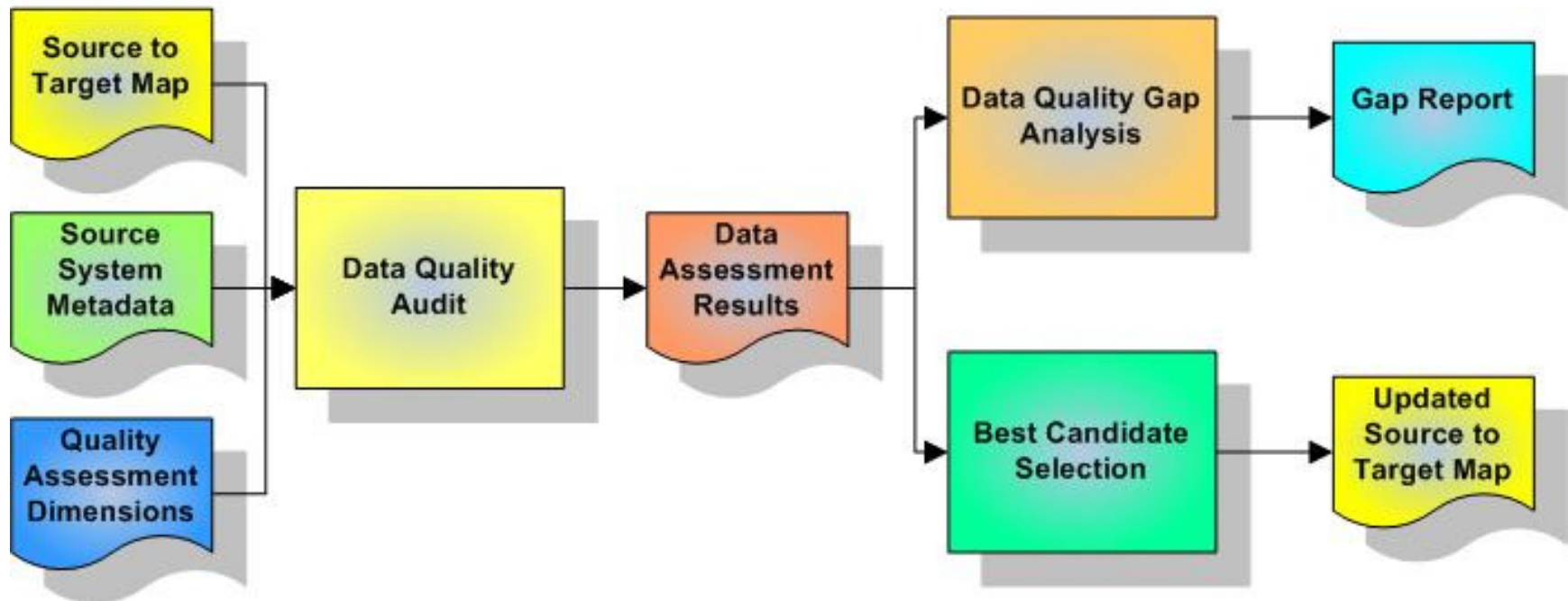
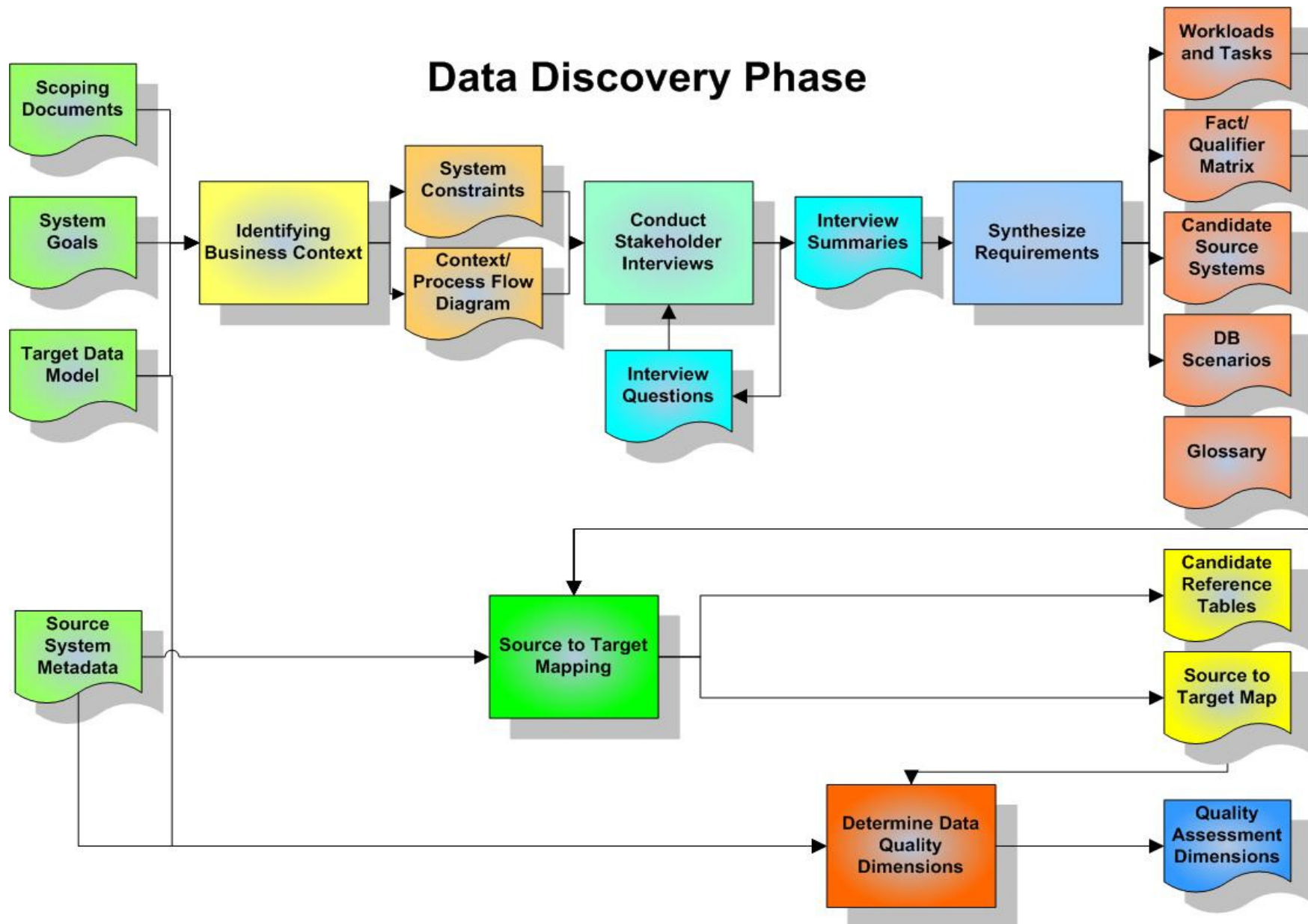# Data Profiling for Data Requirements Analysis

- 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

Target Data - Data Element Name	Source Data									Requirements ID
	System	Table	Column	Data Type	NULL Option	% Null or missing	Min Value	Max Value	Comments	

Data Discovery Phase

Scoping Documents

System Goals

Target Data Model

Identifying Business Context

System Constraints

Context/ Process Flow Diagram

Conduct Stakeholder Interviews

Interview Questions

Interview Summaries

Synthesize Requirements

Workloads and Tasks

Fact/ Qualifier Matrix

Candidate Source Systems

DB Scenarios

Glossary

Source System Metadata

Source to Target Mapping

Candidate Reference Tables

Source to Target Map

Determine Data Quality Dimensions

Quality Assessment Dimensions

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

Questions?

- If you have questions, comments, or suggestions, please contact me

  David Loshin

  301-754-6350

  loshin@knowledge-integrity.com