Whitemarsh
Information Systems Corporation

Data Management Program:
Quality Assessments

Whitemarsh Information Systems Corporation
2008 Althea Lane
Bowie, Maryland 20716
Tele: 301-249-1142
Email: Whitemafsh@wiscorp.com
Web: www.wiscorp.com

Topics

- Net Centric Data Goals

- Data Management and Net Centric Goals

- Data Quality Does Matter

- Key Process Areas in Data Management

- SEI Levels

- Cross Relationship between Net Centric Data Goals and Data Management Assessment

- Data Management Assessment Process

- Benefits and ROI

DoD Net Centric Data Goals	
Goal	**Description**
Make data visible	Users and applications can discover the existence of data assets through catalogs, registries, and other search services.
Make data accessible	Users and applications post data to a "shared space." Data assets, are made available to any user or application under security, etc.
Institutionalize data management	Data approaches are incorporated into Department processes and practices.
Enable data to be understood	Users and applications can comprehend the data, both structurally and semantically.
Enable data to be trusted	Users and applications can determine and assess the authority of the source because the pedigree, security level, and access control.
Support data interoperability	Many-to-many exchanges of data occur between systems, through metadata enhanced.
Be responsive to user needs	Perspectives of users are incorporated to ensure satisfaction.

Understanding Based Data Interoperability	
DoD NC Data Goal	**Data Management Support**
Make data visible	Identifies data assets within natural contexts of mission, organization and function. Standardizes taxonomies, ontologies and classification schemes to view data semantics including where and how implemented
Make data accessible	Includes Discovery Metadata definition in every data asset. Standardized names, definitions and structures
Institutionalize data management	Standardizes strategies for data definition. Multiple layers to ensure define once use many times. Use of ISO standards 11179 for data elements and SQL for data models.
Enable data to be understood	Standardizes vocabularies, commonly inherited semantics, commonly used data model templates, automatic names and definitions based on well defined words. Automatic abbreviations where necessary.

Understanding Based Data Interoperability	
DoD NC Data Goal	**Data Management Support**
Enable data to be trusted	Contains consistent semantics, standard reference tables, completely mapped data models across multiple levels of abstraction.
Support data interoperability	Standardizes data structures, well engineered data transactions, automatic XML wrapping of data, supported by accessible data definitions and contexts.
Be responsive to user needs	Supports reuse of already defined data assets metadata, central knowledge of all data assets and distributed access to same.

From Larry English (www.infoimpact.com)

- 75% of companies report significant problems due to defective data

- 92% of claims Medicare paid to community health centers over one year's time were "improper or highly questionable"

- Wrong price data in retail databases costs consumers as much as $2.5 billion in overcharges annually

- 96,000 IRS tax refund checks were returned as undeliverable one year

Financial Benefits from Data Standardization Data

Activity	Quantity	Cost via technique employed for definition
Starting quantity of columns/fields	19,000	$6.75 million
Elimination of closely named columns and fields reduced the quantity to	3,000	$1.06 million
Elimination of same concept but very differently named columns and fields reduced the quantity to	560	$200,000

Metadata Repository Architecture			
Metadata Levels	**Type and Instance Pairs**		
Fundamental Level			Repository Definition Schema
Repository Definition Level		Repository Schema	Repository Definition Database
Repository Level	Application Schema	Repository Database	
Application Level	Application Database		
	Application Level Pair (ALP)	**Repository Level Pair (RLP)**	**Repository Definition Pair (RDP)**

- Stove Pipe Systems exist if there's one-off **Application Level Pairs**.

- Stove Pipe Systems are prevented when there are **Repository Level Pairs** from which Application Level Pair Systems are based

- Metadata repositories are build through pre accomplished **Repository Definition Pairs**.

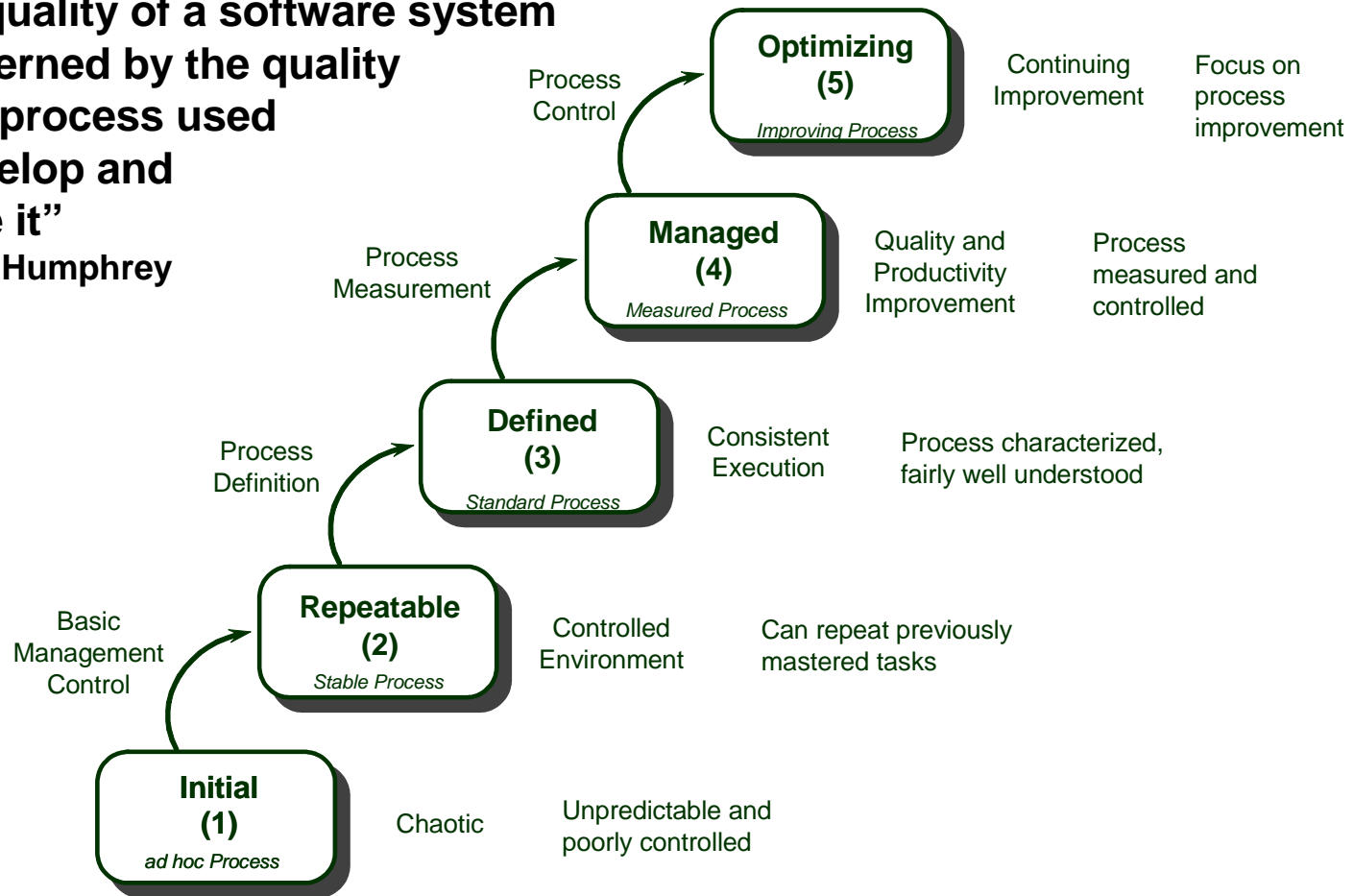## 1995 USAF IRM Weaknesses were Caused from Lack of Metadata Repository

- Current Air Force IRM program is poorly defined with conflicting policy and procedures (*Repository Level Pairs*)

- Technology solutions implemented prior to process improvements (*Application Level Pairs*)

- Current IRM program does not measure customer needs well (*Application Level Pairs*)

- Benefits from sound investments in IRM need to be articulated (*Repository Level Pairs*)

- Lack of standards for creation, collection, accessibility, storage, retrieval, protection, and destruction of electronic information (*Repository Level Pairs*)

- Many stove-piped and incompatible systems require costly changes to share information and to interoperate (*Application Level Pairs*)

- Systems and communication infrastructure based on old costly and inefficient technologies which impede the decisions maker's access to information (*Application Level Pairs*)

- Lack of data standardization which results in inconsistent data, difficult integration, and costly software development and maintenance (*Repository Level Pairs*).
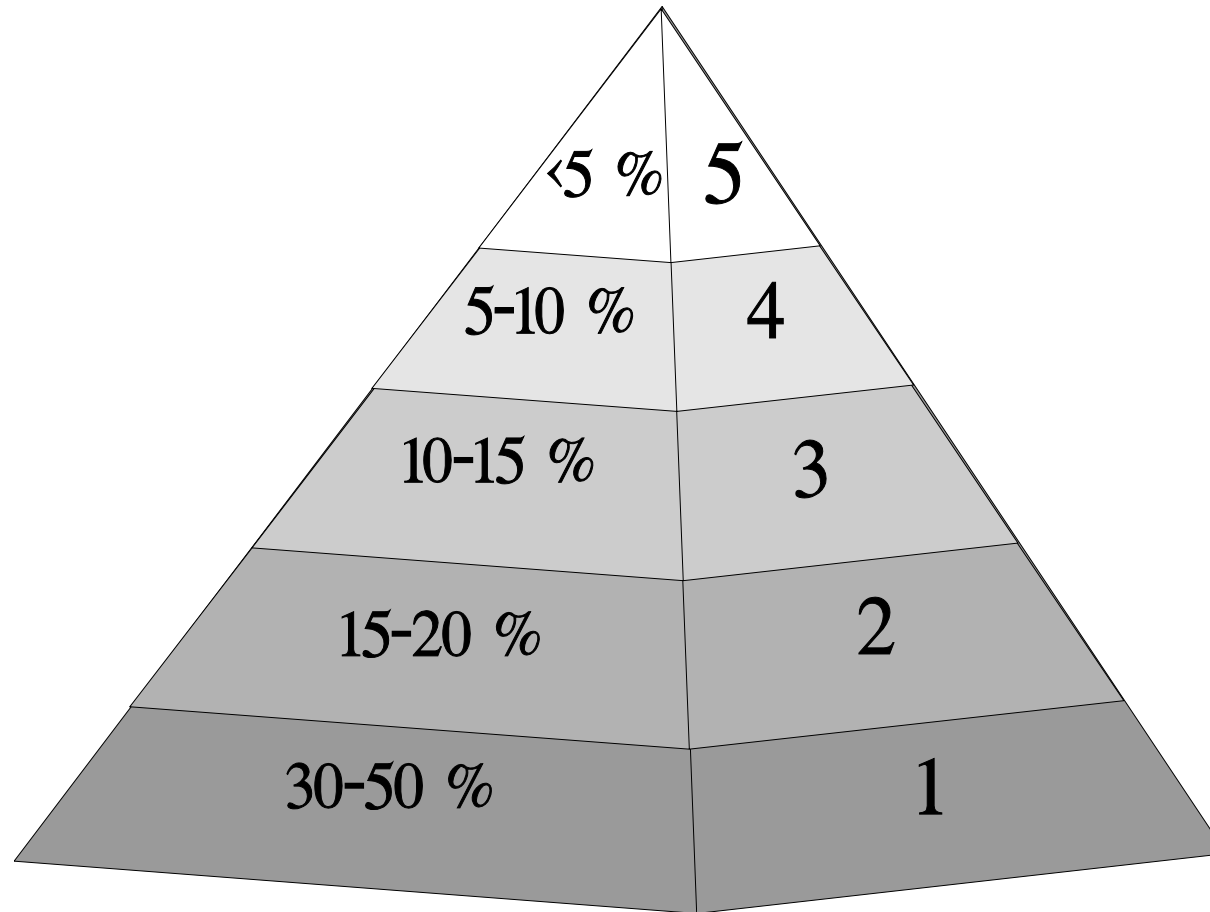
Software Engineering Institute's Increasing Levels of Quality

**"The quality of a software system
is governed by the quality
of the process used
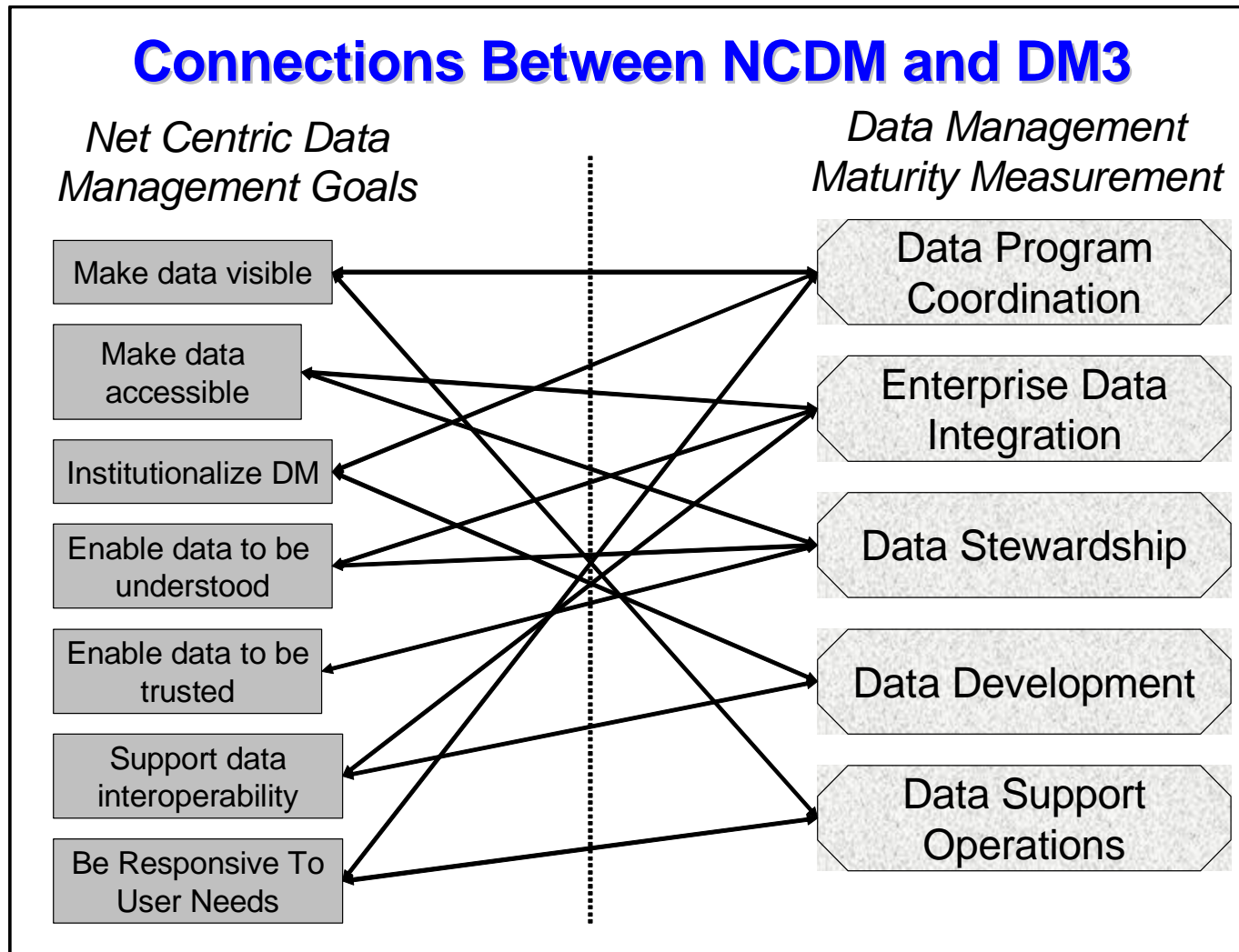to develop and
evolve it"
– Watts Humphrey**

Process
Control

**Optimizing
(5)**
Improving Process

Continuing
Improvement

Focus on
process
improvement

Process
Measurement

**Managed
(4)**
Measured Process

Quality and
Productivity
Improvement

Process
measured and
controlled

Process
Definition

**Defined
(3)**
Standard Process

Consistent
Execution

Process characterized,
fairly well understood

Basic
Management
Control

**Repeatable
(2)**
Stable Process

Controlled
Environment

Can repeat previously
mastered tasks

**Initial
(1)**
ad hoc Process

Chaotic

Unpredictable and
poorly controlled

Distribution of Levels

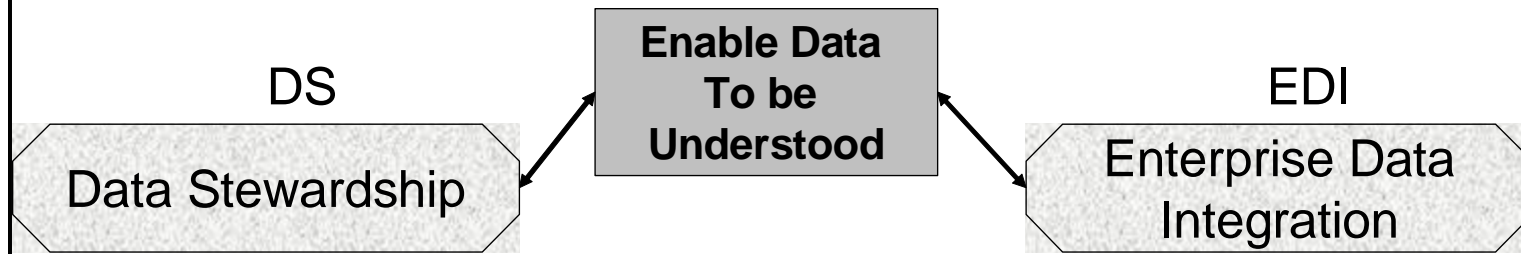# Key Data Management Functions (Parker, Allen, Aiken)

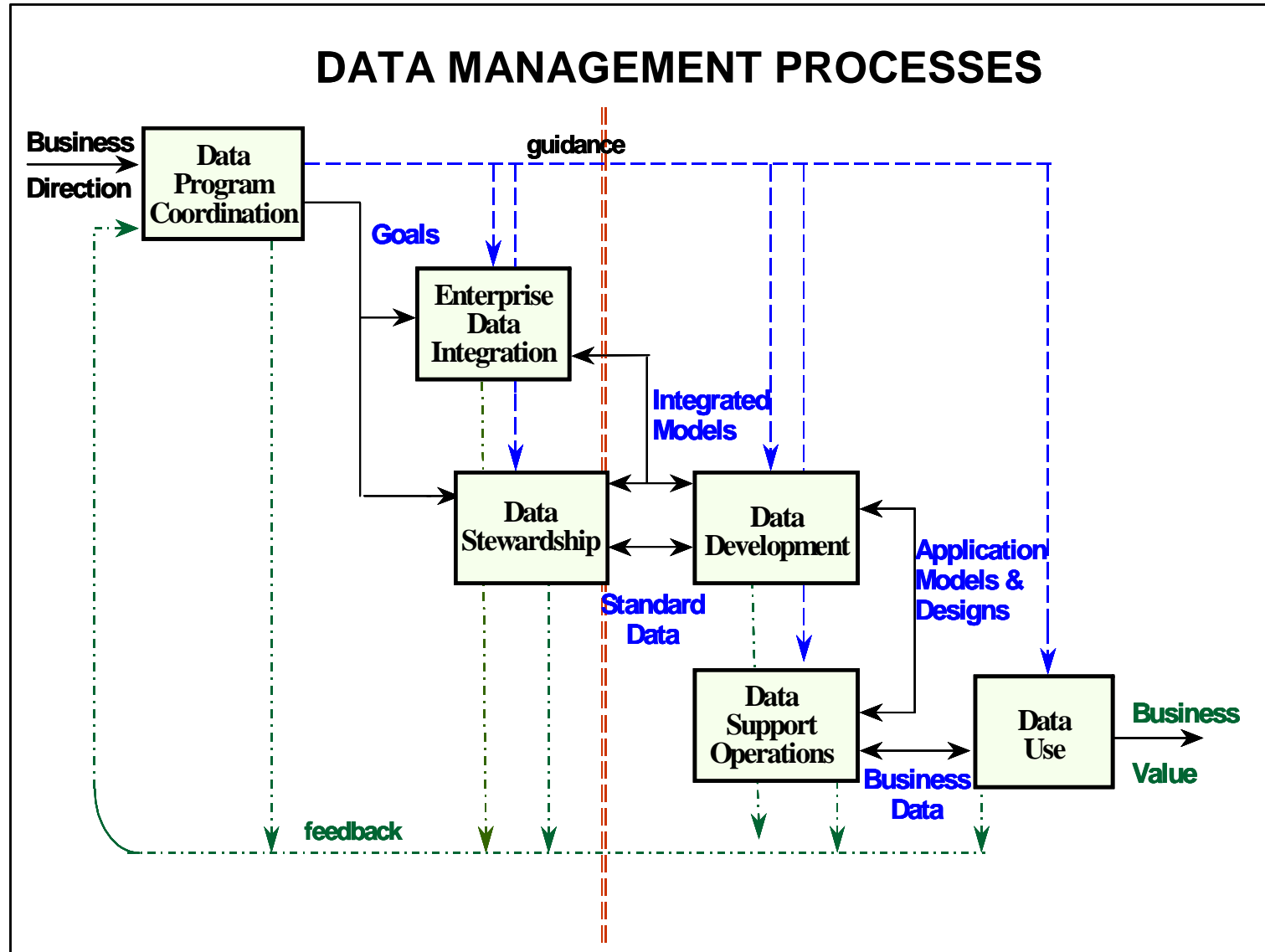| Data Management Process Areas	Description
Data Program Coordination	Management of an enterprise's overall "data program." Includes goals, policies, standards, budgets, etc.
Enterprise Data Integration	Management of all "data specifications." I.e., metadata stored in metadata repositories manifest through data models, business rules, etc. within and across projects, programs, and the enterprise as a whole.
Data Stewardship	Management of the administrative factors (positions, staff, processes, and quality assessments) dealing with "real data" and metadata about the "real data."
Data Development	Activities associated with the creation, use, management and evolution of all "data specifications."
Data Support Operations	Establishment and management of standards and procedures associated with "real data" collection, use, quality and optimization.

Interrelationships among Data Management Processes

Example Mapping: Enable Data to be Understood

```
              Enable Data
     DS         To be                    EDI
            Understood
  [Data Stewardship]              [Enterprise Data
                                    Integration]
```
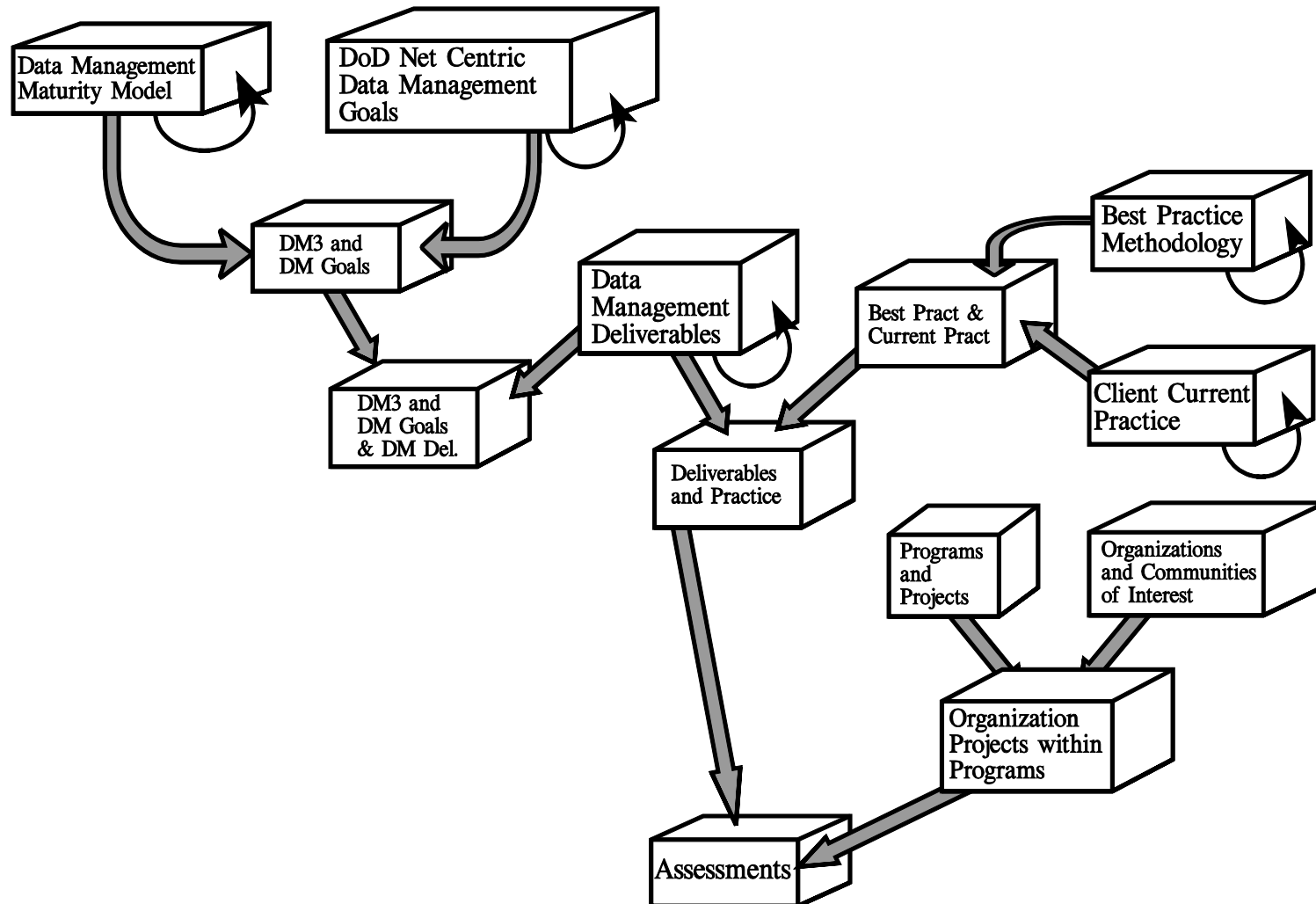
- Understanding is an issue of mapping between syntax and semantics

- Data Stewards (via DS) have access to syntax and semantics of their systems.

- EDI has access to information across systems, and are keepers (with DPC) of standard vocabularies, common templates.
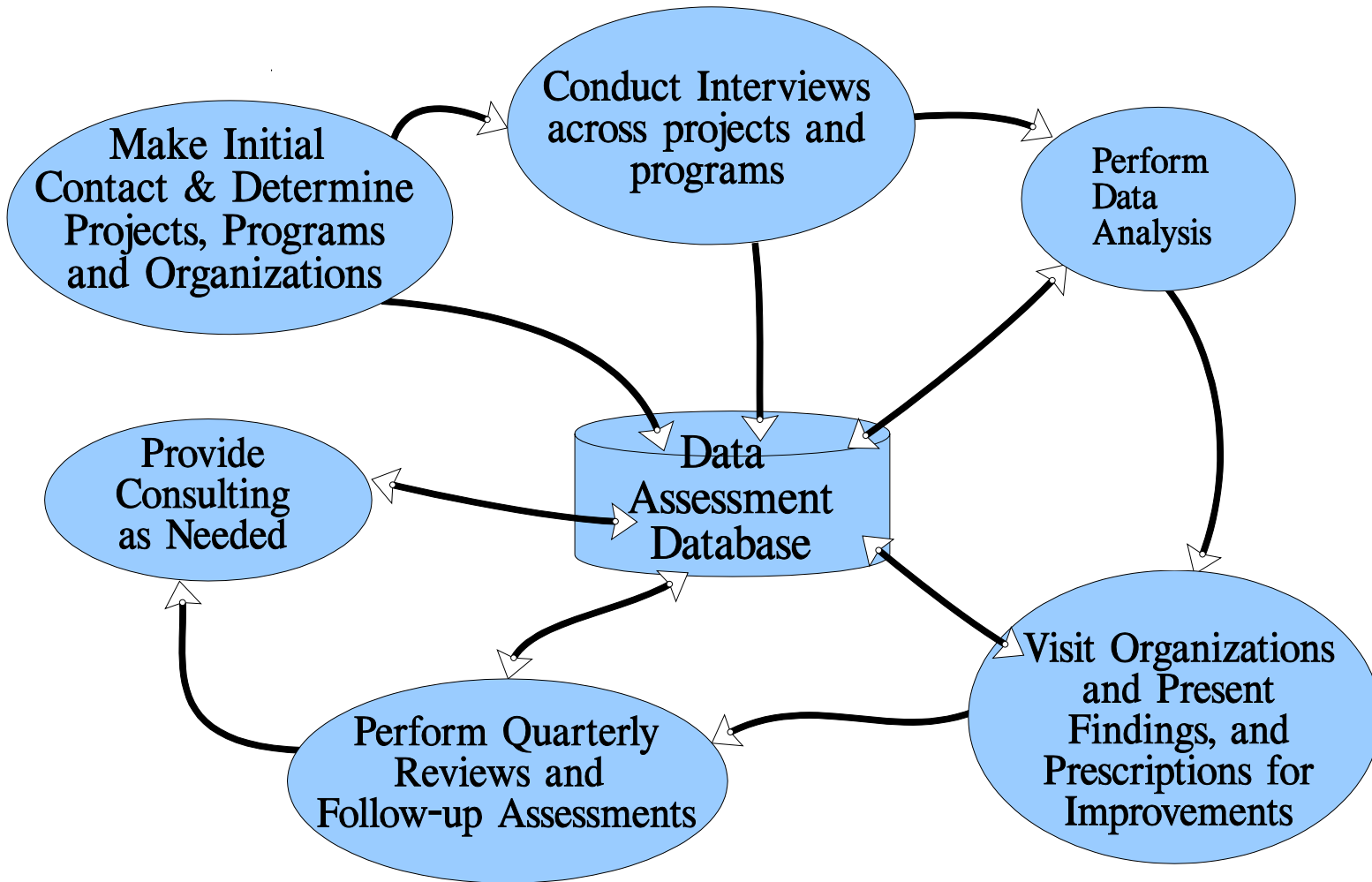
DATA MANAGEMENT PROCESSES

Business Direction →

Data Program Coordination

guidance

Goals

Enterprise Data Integration

Integrated Models

Data Stewardship

Data Development

Standard Data

Application Models & Designs

Data Support Operations

Data Use → **Business Value**

Business Data

feedback

Hi-Level Data Model for Data Management Assessments

Data Management Assessment Process Model

ROI: Solution Benefits

Benefit	Timeframe

• Determines the degree to which a Program/COI/Project data management practice is Net Centric
• Measures Net Centric conformance
    • Repeatable (show incremental progress)
    • Reliable (provides objective results)
    • Independent (all projects can be assessed)
    • Comparable (similar result types are across programs)
• Improves understanding of risk to projects, corrective actions needed.
	Now

• Allows accurate budgeting taking into account the current state, current needs.	**Short Term**
• Finds weaknesses, helps guide investments to improve future state, meeting future needs.	**Long Term**

17