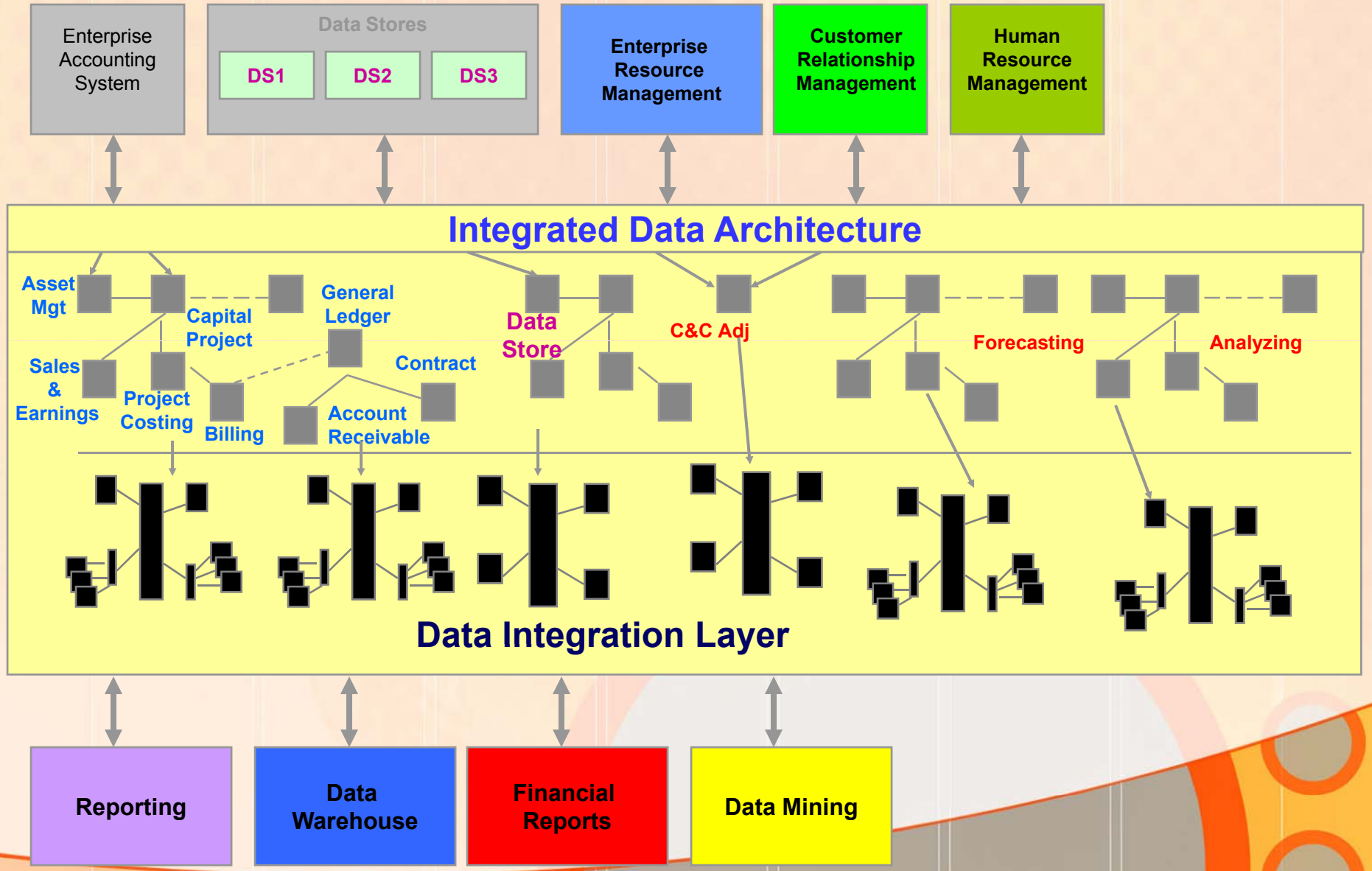
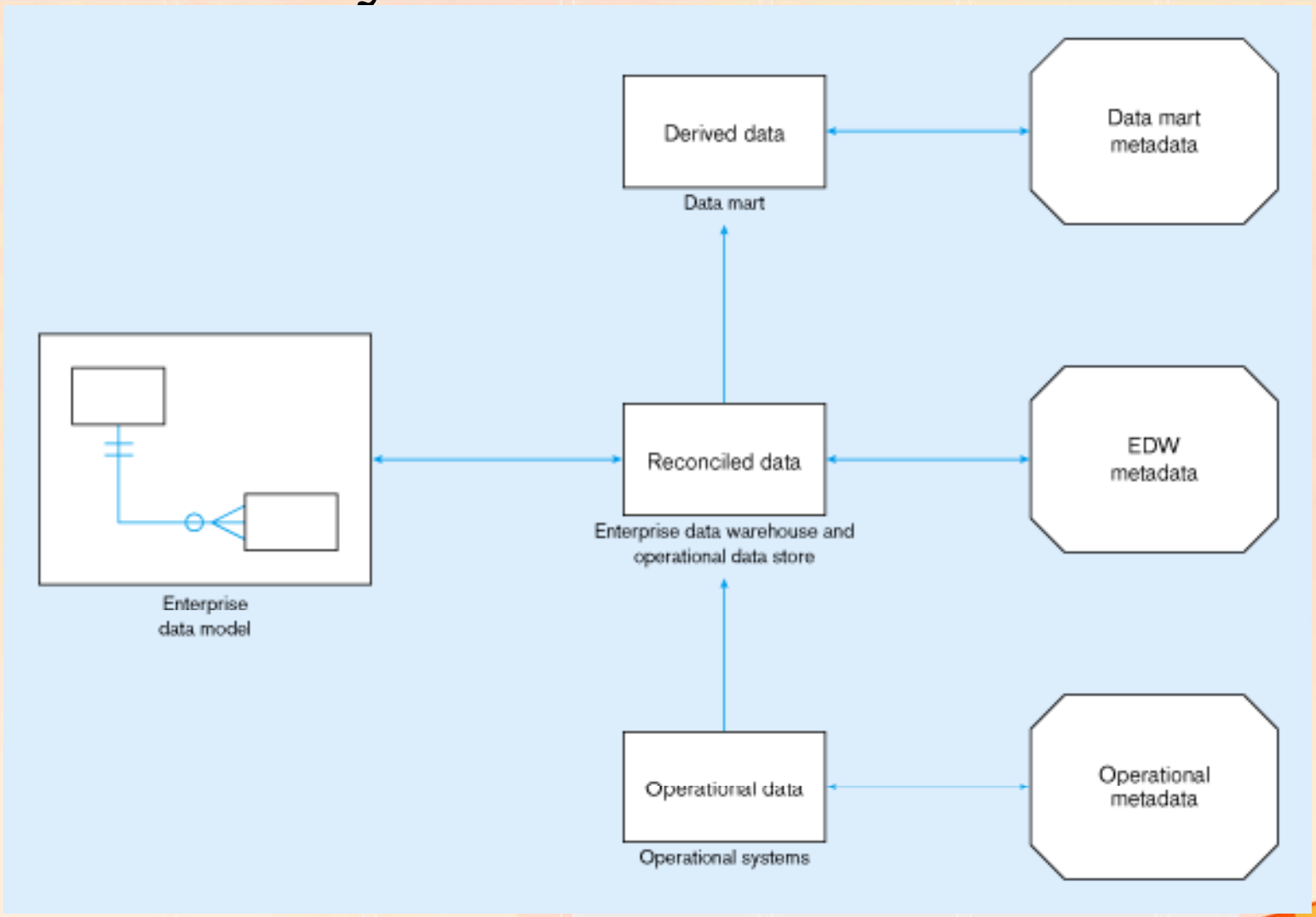
Data Integration

Arman Kanooni

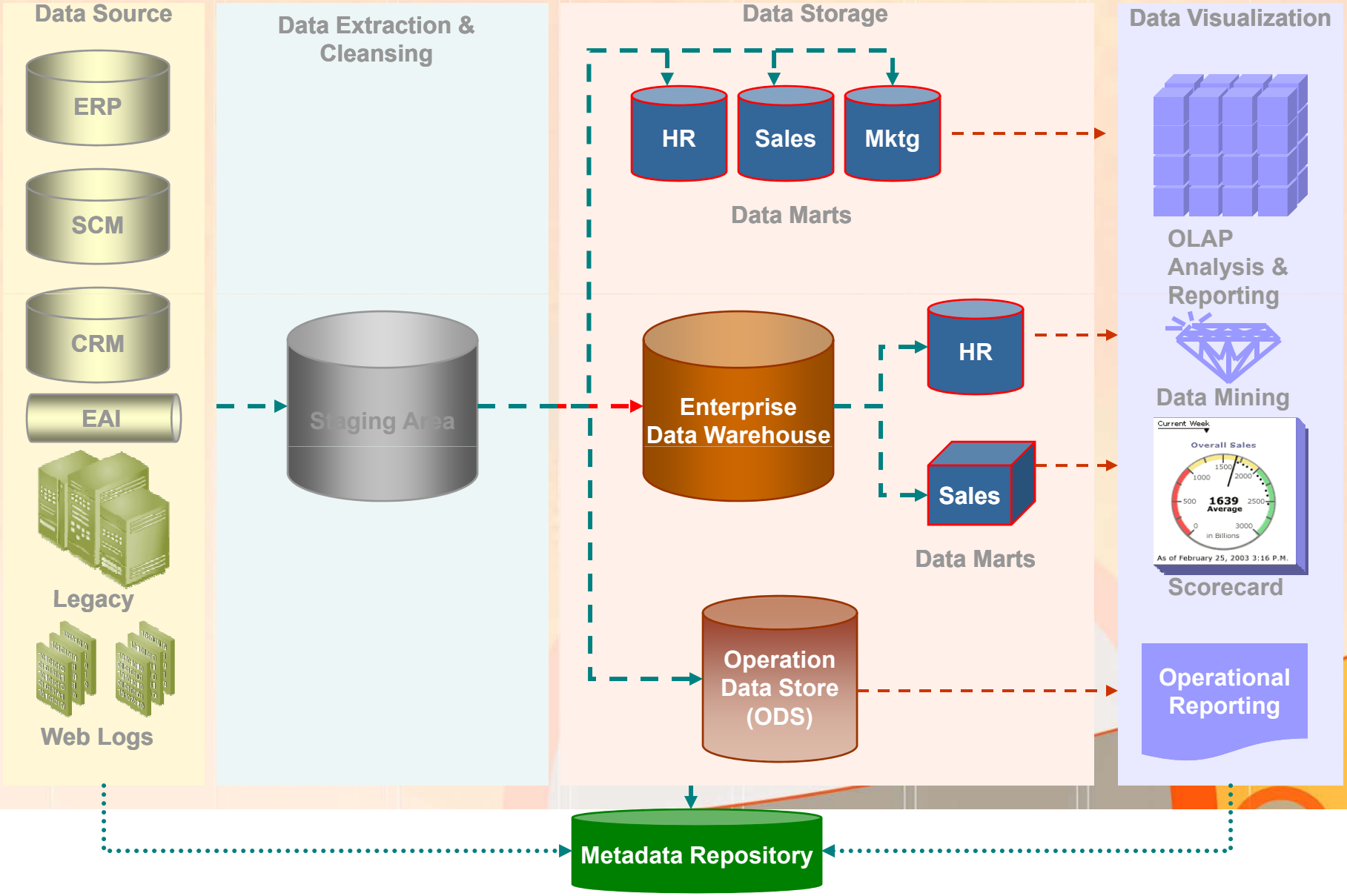
Data Architecture Integration



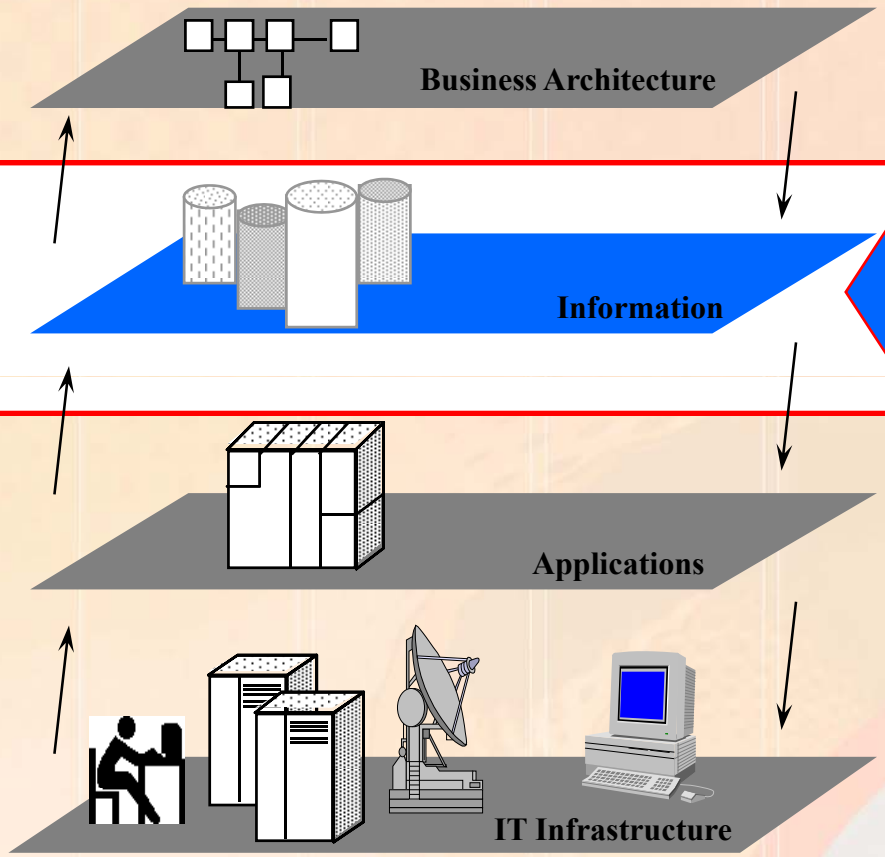
Three-layer architecture



Typical BI/DW Framework



Enterprise Information Architecture



- **Business Processes** reflect a view of what the business must do to be successful
- ***Information represents what must be "known" to effectively execute and support the business***
- **Applications** serve to support business functionality and deliver information to support processes
- **IT Infrastructure** enables access to information and execution of activities by providing the environment and platforms for applications

The are several challenges driving the complexity of large company information architectures.

Information Management Challenges

- Packaged software comes with physical database designs and data update assumptions which may be inconsistent across vendors and business areas
- Rapidly growing complexity of information management
  - Fast moving technologies
  - Globalization of business partners
  - Continued expansion internationally
- Client/server implementations distributed data and processes across platforms (e.g. mainframe, mid-range and distributed servers)
- Security and integration of information is not centrally managed
- Decentralization of application development staff within functions/departments
- Hard to get business function participation and commitment
- Not perceived and supported as a critical path requirement to delivering applications

Typical Information Management Implications

- Expanding roles for the CIO office
  - Coordinating enterprise-wide IT planning and infrastructure
  - Understanding evolving technologies and business implications
- Significant upgrade of IT skills
  - Architecture planning skills
  - Information management and data modeling skills
- New IT processes
  - Define and maintain enterprise standards for Information Management
  - Define, implement, and support an enterprise technical architecture
  - Governance, Prioritization, IT Planning, and Communications
- Systems and Information
  - Increased importance of information integrity internally and externally with customers and suppliers
  - eBusiness demands to interact seamlessly
  - Need to provide enterprise wide access to data on a timely basis
  - Demand for data consistency, standards and integrity

Managing Enterprise Information across organizational boundaries will improve the definition, consistency and sharing of key information.

	Company-Wide Organizations											
Information Subject Area	*Sales*	*Corporate Marketing*	*Cto Office*	*Product Development*	*Product Marketing*	*Operations*	*Local-ization*	*Nts Support*	*Education*	*Finance*	*Legal*	*Hr*
Profit and Loss Statements	X	X		X	X	X	X	X	X	X		
Expenditure Transaction	X	X	X	X	X	X	X	X	X	X		X
Revenue Transaction	X	X		X	X	X	X	X	X	X		
Unmet Customer Need	X	X	X	X	X	X	X	X	X	X		
Customer Profile	X	X	X	X	X	X	X	X	X	X		
Customer Product Feedback	X	X	X	X	X	X	X	X	X	X		
Sales Lead	X				X			X	X			
Marketing Program/Promotion	X	X	X	X	X	X		X	X			
Business Unit	X		X	X	X	X	X	X	X	X		
Product Line	X	X	X	X	X	X	X	X	X	X		
Product	X	X	X	X	X	X	X	X	X	X		
Product Defect (Bug)	X			X	X	X	X	X	X			
Bill of Materials	X	X	X	X	X	X	X	X	X			
Product Specification	X	X	X	X	X	X	X	X	X		X	
Product Registration	X	X		X	X	X	X	X	X		X	
Distributor	X	X		X	X	X	X	X	X		X	
Reseller	X	X		X	X	X	X	X	X		X	
Original Equipment Manufacturer	X	X	X	X	X	X	X	X	X		X	
Business Partner	X	X	X	X	X	X	X	X	X	X	X	
Employee	X	X	X	X	X	X	X	X	X	X	X	X
Process Correction Information	X	X	X	X	X	X	X	X	X	X	X	X

To improve information architecture, Company will need to focus on both logical and physical design improvements.

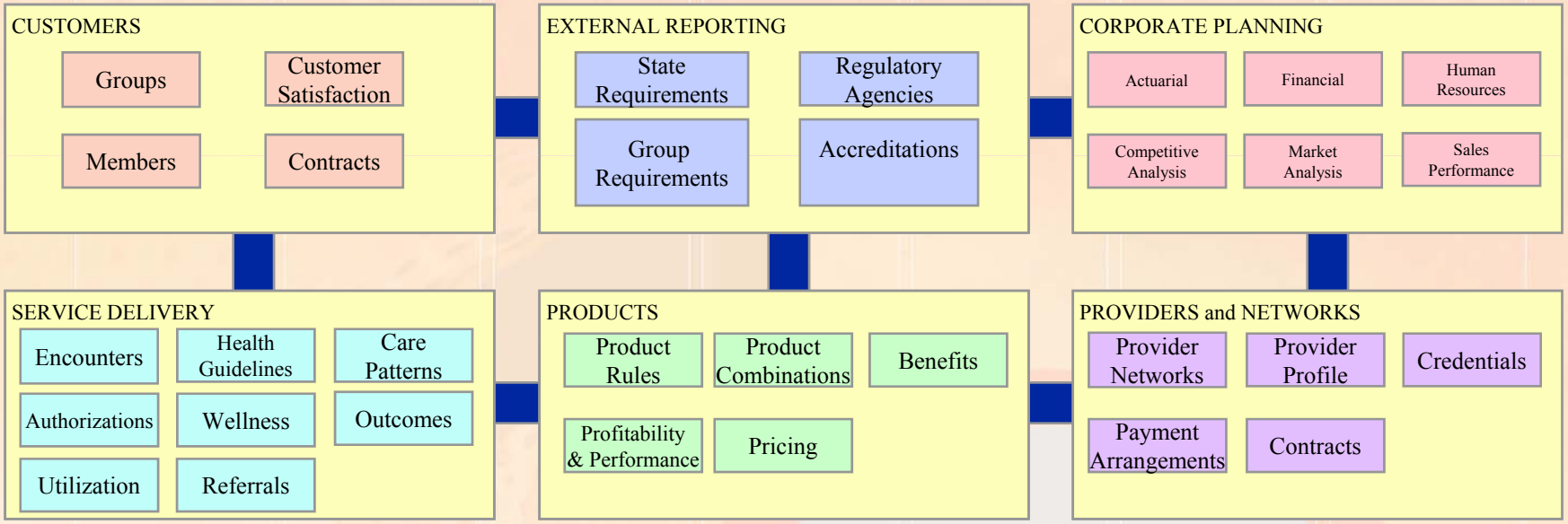
- Logical Design -

- *Logical models*
  - *Enterprise Model*
  - *Business Area Models*
  - *Application Data Models*
- *Data Subject Areas*
- *Entities*
- *Attributes*
- *Entity Relationships*
- *Cardinality*
- *Typically third normal form*

- Physical Design -

- *Data bases*
- *Data access*
- *Physical storage*
- *Data distribution*
- *Data Schemas*
- *Tables*
- *Fields*
- *Table relationships (Foreign Keys)*
- *Denormalized for physical performance*

Example Logical Enterprise Data Model



Enterprise Information Framework

Categories	Transactional or Operational Data	Operational Data Store	Information Warehouse (Detailed, Summary, Meta, Ref.)	Data Mart (Highly Summarized, Department)
	Ex: Baan Master Files Baan Inventory Item	Ex: Inventory History (History Files)	Ex: BCA Example?	Ex: Inventory Turns versus the competition
Characteristics	• Good performance on-line • Current in time	• Consistent and integrated • Near current	• Enterprise subject data • Daily, weekly, monthy	• Specific
Application	Reporting	Decision support and query	On-line analytical processing (OLAP)	Enterprise & Department OLAP

— Information Framework —

	Generate Demand and Sell	Make Product	Manage Finances	Manage People
I. Strategic Planning	**Long-Range Market Planning** ■ Long-Range Marketing Strategies &Forecasts ■ Modeling of Alternative Market &Distribution Strategies ■ Modeling of Long-Range Sales ■ Contract Pricing & Term Strategies	**Business Development** ■ Monitor Competitor activities (e.g. worldwide price movements) ■ External Financial Analyst opinions ■ Industry Analyst information ■ Key Development Performance Indicators ■ Enterprise planning ■ Performance Trends ■ Emerging markets ■ Strategic Partnerships and alliances	**Long-Range Financial Planning** ■ Analysis of Investment Alternatives ■ Performance trends of business units ■ Product line and product trends ■ Financial & Accounting Policy ■ Financial & Business Planning Models	**Long-Range Personnel Planning** ■ Long-Range Productivity ■ Improvement Strategies ■ Competency and workforce trends ■ Personnel Policies ■ Union & Government Relations
II. Operational Planning	**Market & Sales Planning** ■ Marketing Planning ■ Pricing Planning ■ Productivity Planning ■ Contract & Customer planning ■ Sales Forecasts ■ Price protection & reserves for returns ■ Distribution Planning ■ Transportation Cost Standards	**Product and Operations Planning** ■ Labor Planning ■ Release management ■ Critical Path analysis results ■ Operations Scheduling ■ 3rd Party Planning ■ Material & Supply Inventory Planning ■ Research & Development Activities ■ Manpower Planning • Project Assignments • Training ■ Maintenance Planning ■ Performance Standards	**Financial Planning** ■ Budget planning by operational structure ■ Financial planning for operational structure ■ Profit & Gain Sharing Planning ■ Balance Sheet Planning ■ Cash Planning ■ Tax Planning - Depletion ■ Depreciation, Investment Credit ■ Royalties Planning ■ Capital Expenditure Planning ■ Equipment, Facilities ■ Financial Performance ■ Standards & Objectives	**Personnel Planning** ■ Wage, Salary & Benefit Planning ■ Career Path Planning ■ Competencies ■ Planning for Upgrade of Target Skills ■ Labor Contract Negotiation ■ Planning & Reporting
III. Operational Reporting	**Sales & Marketing Results** ■ Contract Negotiation & Re-negotiation ■ Contract Administration & Escalation ■ Contract Variance & Analysis • Shipments • Prices • Quality • Profitability ■ Sales Analysis ■ Customer/Reseller/Partner Profiles ■ Sales Forecast Performance ■ Spending vs. Budget ■ Customer Service Performance ■ Carrier Performance & Cost	**Operations Results** ■ Sales by business unit, product line, product ■ Daily Cost & Statistical Performance Reporting (on demand) ■ Reporting (Actual vs. Standard) ■ Delay & Major legal events ■ Personnel Utilization & Performance ■ Labor Utilization & Performance ■ Process Metrics ■ Purchasing Performance & Vendor Performance **Product Development** ■ Profitability (project, business unit, product line) ■ Project Productivity Reporting ■ Release Management Schedule ■ Program/dependencies & critical path ■ Project Scheduling & Status Reporting ■ Estimating & Cost Control ■ Commitment Accounting ■ Product Cash Flow Projections	**Financial Control** ■ Profitability Reporting by business Units, Product Lines, Products, Geography and Vertical Industry Segments ■ Budgetary Control & Variance by operational structure (business units) ■ Responsibility Reporting ■ Project Control ■ Sales Contract Profitability Reporting ■ Projection and Forecasting Capability ■ Sales Contract Estimating & Cost ■ Financial Statements ■ Funds Analysis ■ Analysis ■ Capital Budget Status	**Personnel Administration** ☹ Attrition Analysis and Trends ☹ Benefits Administration ☹ Education & Training History ☹ MBO Reporting: Hoshins ☹ Salary , Compensation & Job History ■ Pension Plan Administration ■ Absenteeism Reporting ■ Legal / Environment Reporting - EEOC, OSHA, Etc. ■ Overtime Control ■ Labor Contract Monitoring
IV. Transaction Processing	**Contract Management** ■ Contracts ■ Terms and Conditions ■ Customer/ Partner ■ Products and Terms **Order Entry & Billing System** ■ Order Entry - Ship To & Product ■ Requirements ■ Control ■ Shipping Control ■ Billing ■ Adjustments ■ Accts. Receivable ■ Credit & Collect. **Distribution System** ■ Shipment ■ Scheduling ■ Product ■ Fleet Status ■ BOM ■ Delivery ■ Reporting & Reconciliation- ■ Transportation- ■ Cost Control	**Production Reporting System** ■ Operational results ■ Actual revenue ■ Actual expenditures ■ Profitability ■ Maintenance Work Orders • Labor Performance • Cost Control **Program & Project Management** ■ Scheduling & Usage Reporting ■ Program & project man days ■ Plan vs. actual by project ■ Analysis of project impacts due to slippage and project dependencies ■ Project resource costs (plan vs. actual) ■ Resource utilization ■ Etc. . . . .. **Procurement** ■ Purchasing - Supplies, ■ Outside Services ■ Material Receiving ■ Material Distribution ■ Usage Reporting ■ Vendor Analysis ■ Inbound Warehousing & Inventory Control	**Financial Systems** ■ Support for Intra Company Transfers ■ Project Expenditure and Revenue Tracking ■ Financial Operational Reporting to business units ■ Cash Management/Accounts Receivable ■ Investment management ■ Accounts Payable ■ General Ledger information ■ External Reporting and SEC filings ■ Chart of accounts design to support appropriate expenditure classification for business units ■ Budgeting ■ Fixed Assets Accounting ■ Reserves & Royalty Accounting ■ Cost Accounting ■ General Accounting - Reserves, ■ Product returns ■ Price protections ■ Cash Accounting	**Human Resource Management Systems** ■ Attrition reasons/exit interview results ■ Skills and competency Inventory ■ Accurate headcount by organization and location ■ Compensation data of base salary, benefits & options ■ Performance results ■ Payrolls ■ Labor Distribution ■ Absence Reporting ■ Work History ■ Training ■ Pension & Benefit Plans ■ Recruiting ■ Applicants

(facer)

The information architecture direction addresses operational and corporate levels of information needs.

Operational
(e.g. Provider or Customer Inquiry Support)

Strategic
(e.g. Product Profitability Analysis)

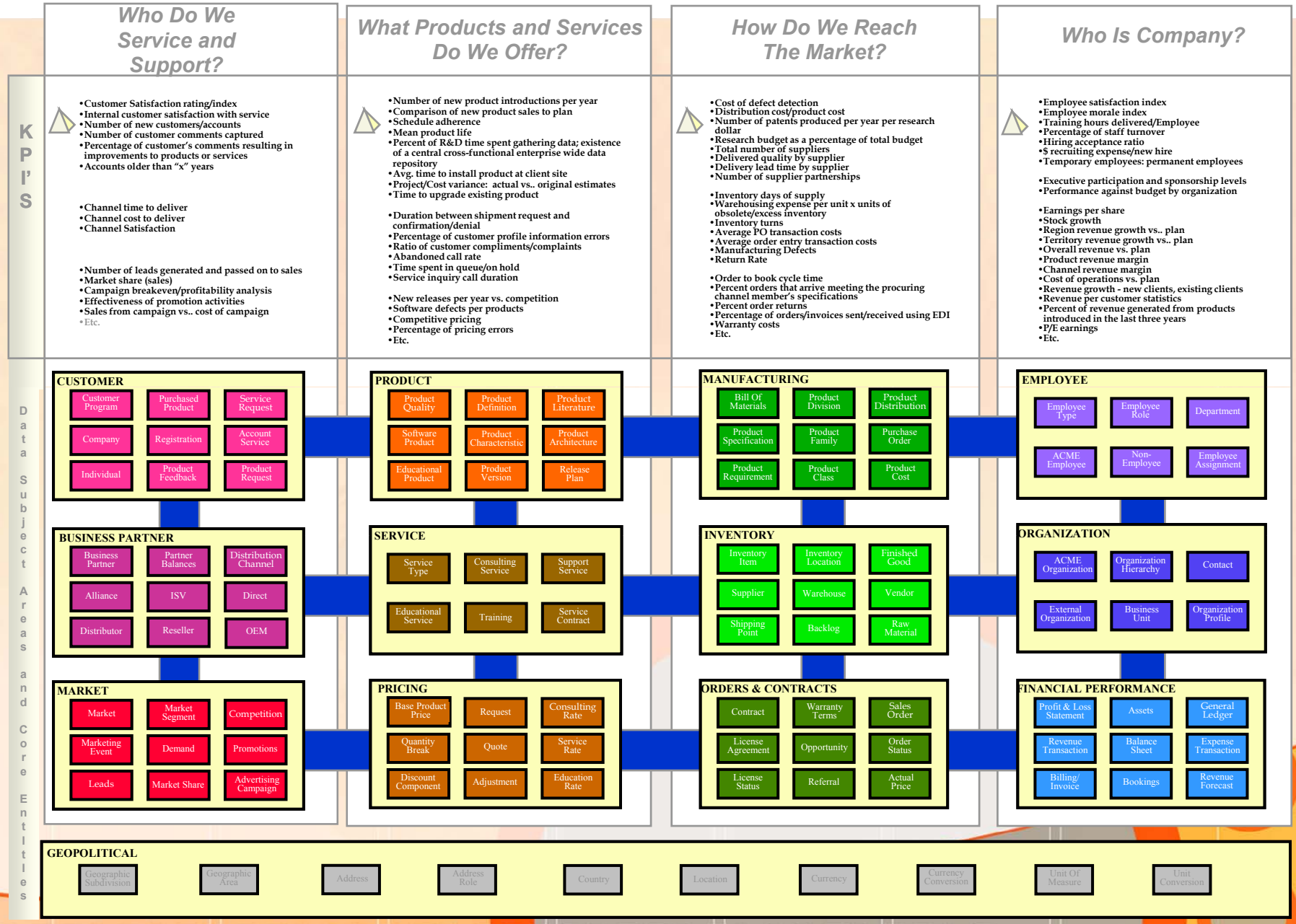
Characterized by:
- *Frequent Warehouse Updates*
- *Increased Data Currency*
- *Less System Stability*
- *Less Reporting Consistency*
- *Less System Availability*

Characterized by:
- *Periodic Warehouse Updates*
- *Decreased Data Currency*
- *Increased System Stability*
- *Increased Reporting Consistency*
- *Greater System Availability*

- *Requires larger number of current data fields to support scope of operational need (e.g. date and time stamps, status fields).*
- *Data needs to be current enough to support operations in progress.*
- *Source system processing needs to be coordinated with repository updates to provide accurate snapshot information from disparate systems.*
- *Infrastructure to support frequent updates is more costly.*

- Requires statistically significant volume of historical data.
- Data currency of 30 days or greater is acceptable and in some cases more desirable due to greater stability of older data.
- Updates from source systems may be scheduled to take advantage of data coordination that occurs as a result of normal processing (e.g. month-end).
- Infrastructure to support monthly updates is less costly.

Logical Enterprise Information Model - *REPRESENTATIVE* -



** Representative Data Subject Areas and Core Entities*

Types of Metadata

- *Business Metadata*
  - *data definitions, rules*
- *Technical Metadata*
  - *ETL, aggregations, derivations, systems arch*
- *Operational Metadata*
  - *schedule runs, data loads, errors*
- *User Management Metadata*
  - *Users, access rights, roles*
- *Reports Metadata*
  - *Report field definitions, scheduling, distribution*
- *Run time Monitoring*
  - *User access statistics, table activity*

Why is Metadata Important?

- Business Benefits:
  - ✔ *Facilitates more informed business decisions*
  - ✔ *Facilitates effective communications*
  - ✔ *Creates efficiencies in doing business (time savings)*
  - ✔ *Captures intellectual capital*
  - ✔ *Higher quality information*

Why is Metadata Important?

Facilitates more informed business decisions

- *Data warehouses/data marts have much less value without meta data*
- *Insight into information assets*
  - *What does total sales **really** mean and how was it derived*
- *Taking actions based upon consistent, enterprise wide, definitive knowledge*
  - *Can have a major positive affect on operations, sales, customer service and other function!*

Why is Metadata Important?

Facilitates Effective Communication

- *Without understanding, there is Confusion*
  - *Data + Metadata = Information*
- *How Can One Effectively Use Data Without Knowing What it is?*
  - *For example, What is sales revenue?*
    - *Amount billed?*
    - *Amount sold?*
    - *Amount collected?*
    - *Amount on general ledger?*
    - *Discounts?*
  - *What is a product?*
    - *Does it include customized solutions configurated for a specific customer?*
    - *Does is include services as well as goods?*

Why is Metadata Important?

Facilitates Work Efficiencies

- *40-90% of many people's time spent gaining access to data and figuring out what it means*
- *Often many people spending time to synchronize data*
- *Shift from discovery mode to analysis mode*
- *More time using data and less time getting it*

Why is Metadata Important?

Preserves Intellectual Capital

- *If data definitions, transformation, derivatives, sources and other knowledge about the data is stored in peoples' heads, then, WHAT HAPPENS IF/WHEN THEY LEAVE?*

Applies to business and technical meta data

Why is Metadata Important?
Higher Quality Information

- *Knowing more about the data*
  - *What is or what is not included*
    - *Returns not included in sales figures*
    - *Certain number of sales records not loaded in DW*
- *Accurate business analysis*
- *Piece of mind/confidence*
- *Regulatory issues*

Why is Metadata Important?

Technical Benefits

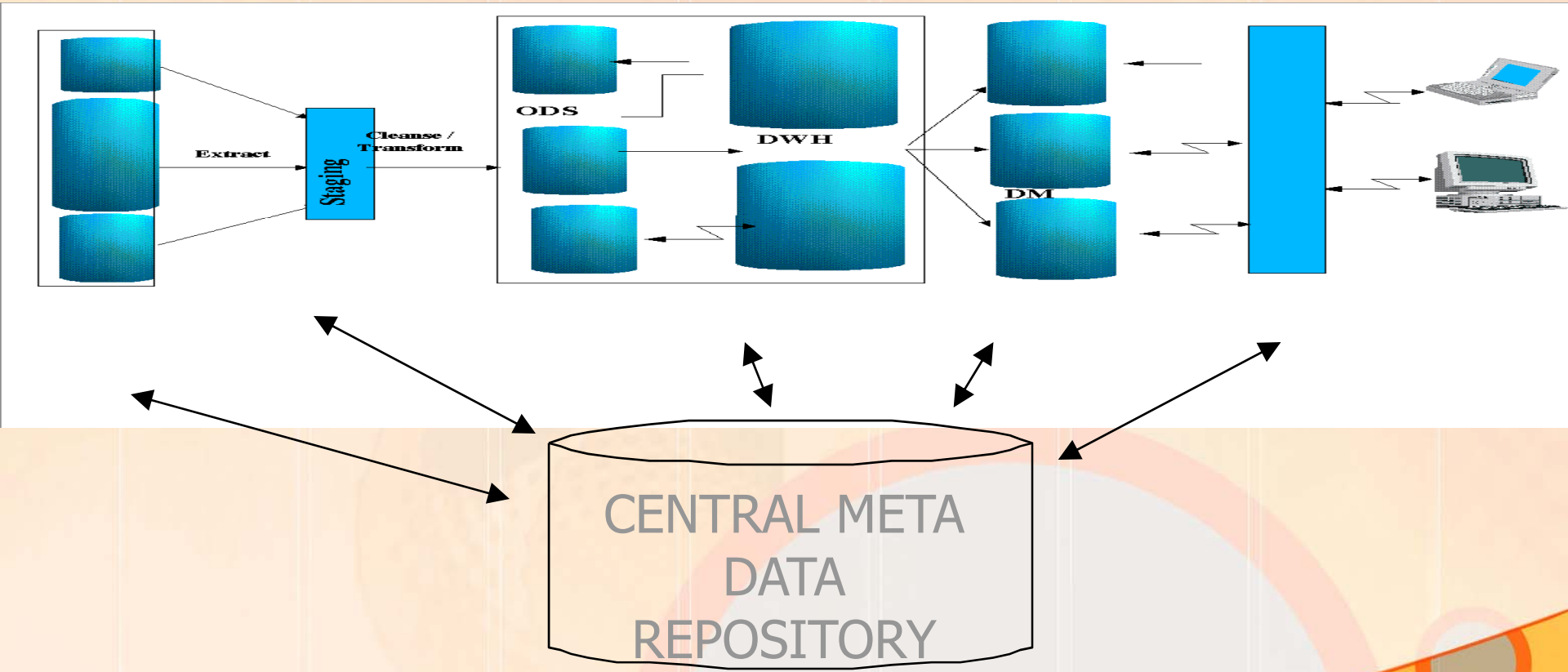
- Less time for IT development (in long run)
  - Reusability and consistency of semantics
  - Less time researching, more time analysing and developing
- More maintainable systems
  - Provides documentation for understanding systems
  - Fewer interfaces
  - Efficient impact analysis and change management
  - Less confusion
- More stable systems (longer life)
  - Shows integration points
  - Foundation for corporate information architecture
- More effective usage of systems
- Basis for effective integration
  - Needed for effective data management
  - Basis for data stewardship
  - Integral part of EII

Metadata Strategy

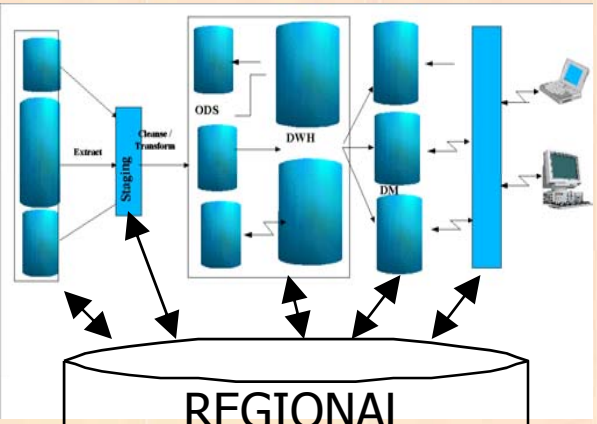
- *Metadata strategy: policy regarding how metadata is managed, controlled and used (from CWM book)*
  - *Who will update metadata?*
  - *Who will access metadata and how?*
  - *Metadata owned by the business, for the business?*
- *Closely linked to data stewardship*
  - *Largest challenge:common understanding of data*
- *What tools will help business manage meta data*
  - *Front end creating/updating/accessing of meta data?*
  - *Version control (to what level?)*
  - *Workflow engine?*
  - *Searching, impact analysis and identify inconsistencies?*

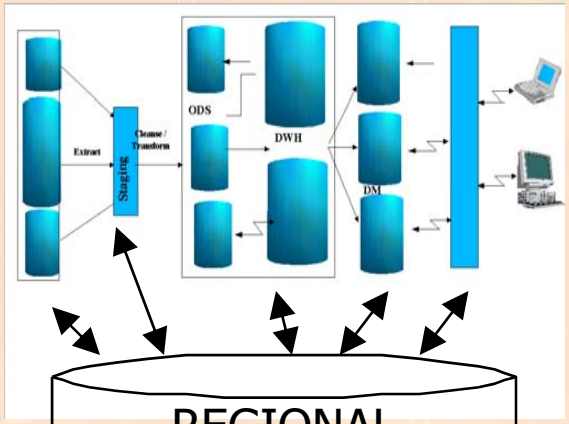
Centralized Metadata Architecture

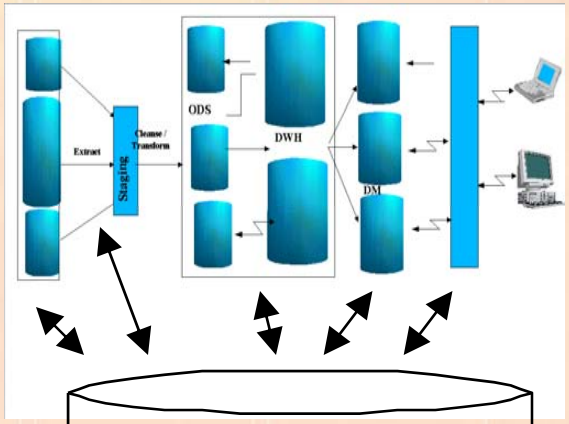
- *All metadata integrated and stored in central data store*



Decentralized Metadata Architecture (Federated)







REGIONAL META DATA REPOSITORY

REGIONAL META DATA REPOSITORY

REGIONAL META DATA REPOSITORY

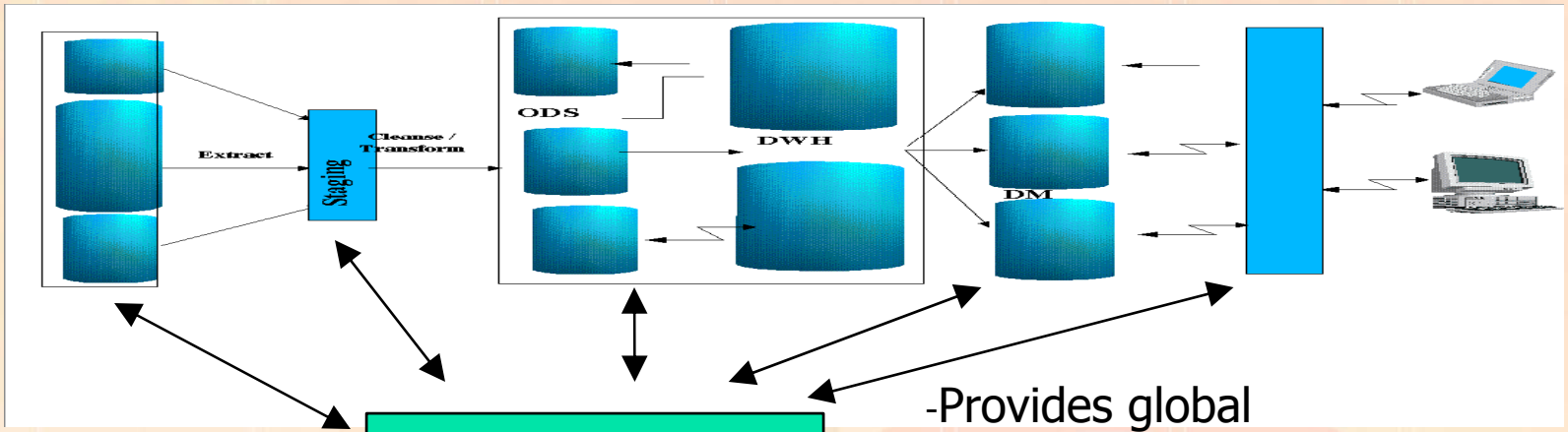
GLOBAL META DATA REPOSITORY

-Mandates consistent meta data schemas

-Provides mechanism for sharing meta data between local meta data repositories

Distributed Metadata Architecture

- *Meta Data is not maintained in the repository*
  - *Only linked to the current meta data stores*
  - *Meta data accessed at time of need*



GLOBAL meta data SOLUTION

-Provides global meta data Access

-No meta data stored here

-only linkages to meta data sources

Define ROI early and Track to ROI

- *Data management should be run like a business within a business – show the value*
- *ROI very difficult to measure later if baselines are not set*
- *ROI realized in short term and long term*
- *Some metrics to measure ROI*
  - *% of time people spend getting information on data*
  - *% of time redoing efforts that have previously been done because no documentation*
  - *% of interactions frustrating customers, employees, and partners because of confusion*
  - *Mistakes caused because of uninformed or misinformed decisions*
  - *Quality and usability of decision support information*

Don’t Boil the Ocean

- *Tremendous amount of metadata*
- *What is most useful? (to start with)*
  - *Definitions*
  - *Source to target mappings*
  - *Derivations*
  - *Processes*
- *For what area of the business? (to start with)*
  - *Customer*
  - *Product*
  - *Sales Analysis*
- *Potential pilots*
  - *Subject data area*
  - *Portion of a data warehouse*

Universal Data Model for Metadata (Conceptual)