Enterprise Data Quality Management for USAF Operations Support

ABSTRACT----------

The USAF, through the Expeditionary Combat Support System (ECSS) Program Office, is piloting an initial capability to manage the quality of data throughout the logistics and other Operations Support domains. This pilot will set up an Enterprise Data Quality Management Service (EDQMS) that builds out existing deployed infrastructure to provide a vendor-neutral data quality metrics database, a data quality metadata exchange (DQME) standard, and set of data quality management tools. The EDQMS will also establish end-to-end DQ lifecycle processes and governance structures. The pilot will then exercise the EDQMS with a set of inventory data from several legacy information manufacturing systems. The overall approach shall be flexible and generic enough for application to other information data products within the Logistics and other AF Operational Support domains.

BIOGRAPHY----------

Christopher J. Sharbaugh
Principal Advisor to Director
US Air Force

Christopher J. Sharbaugh, a member of the Senior Executive Service, is Principal Advisor to Director of Transformation DCS/Logistics, Installations & Mission Support, WPAFB, Ohio. He is responsible for integration of enterprise-level data in support of the Expeditionary Combat Support System (ECSS). He monitors legacy systems; plans, organizes, and evaluates data integration; reviews all modernization efforts; and develops and provides guidance on data integration to the logistics community and other AF functionals, Services and Agencies. Mr. Sharbaugh began his career as a researcher for the DLA/DTIC Crew Systems Ergonomics Information Analysis Center. He has held a series of positions at KPMG LLP/BearingPoint Consulting and Morgan Borszcz Consulting supporting various AF data initiatives. As an Enterprise Data Architect at General Electric, Mr. Sharbaugh performed an integration function, developing and implementing strategies of data quality, ERP data migration, master data management, data integration, and enterprise reporting across several lines of business.

David K. Becker
Principal Information Systems Engineer
The MITRE Corporation

David Becker is a Principal Information Systems Engineer with the MITRE Corporation. He works out of the Dayton, OH site at Wright-Patt AFB as chief architect of AFMC/ESC's 554 Electronic System's Group (554 ELSG). He is currently working on a number of projects in enterprise architecture, information quality, data strategy, and program acquisition. David has over 30 years of experience in software development and information technology. While working at Lexis-Nexis and CSC, he has had a broad range of assignments, including senior level information technology and business consulting, technical leadership and management, project management, product research & development, seminar and workshop development, college level

computer science course development and instruction, industrial liaison, international standards development, systems administration, and systems analysis, design and implementation. David's particular areas of strength include business, application, data and technology architectures, systems dynamics, project management, statistical process control, information search and retrieval, and artificial intelligence.

United States Air Force

Integrity - Service - Excellence

Enterprise Data Quality Management for USAF Operations Support

U.S. AIR FORCE

July 15-17, 2009

Chris Sharbaugh
USAF SES

Dave Becker
The MITRE Corporation

1

U.S. AIR FORCE

Abstract

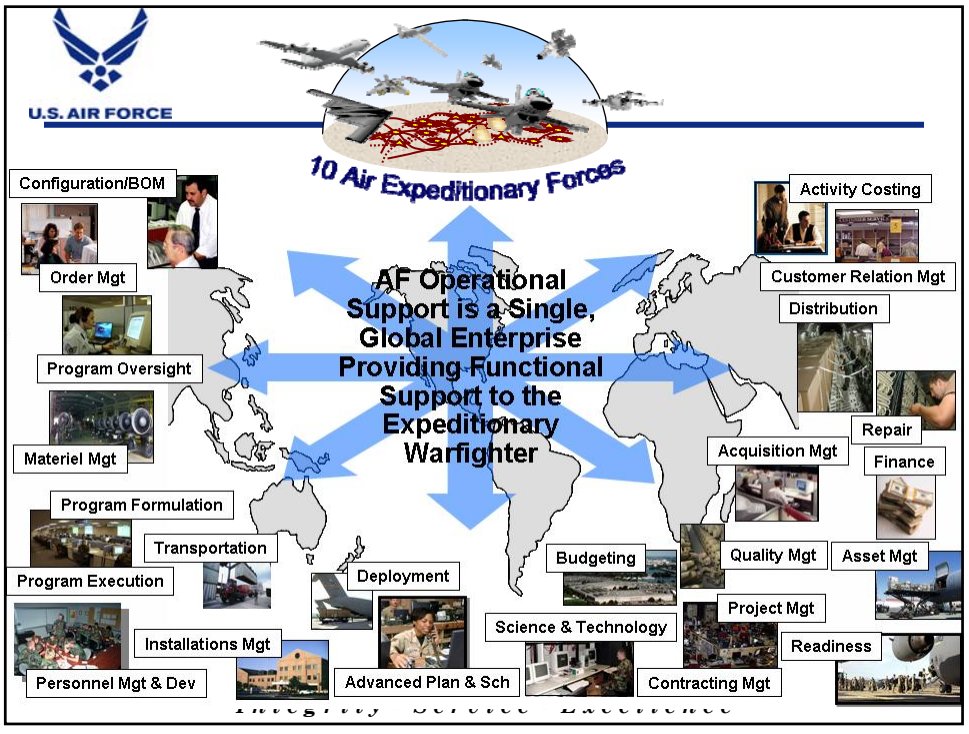
Enterprise Data Quality Management for USAF Operations Support

The USAF, through the Expeditionary Combat Support System (ECSS) Program Office, is piloting an initial capability to manage the quality of data throughout the logistics and other Operations Support domains. This pilot will set up an Enterprise Data Quality Management Service (EDQMS) that builds out existing deployed infrastructure to provide a vendor-neutral data quality metrics database, a data quality metadata exchange (DQME) standard, and set of data quality management tools. The EDQMS will also establish end-to-end DQ lifecycle processes and governance structures. The pilot will then exercise the EDQMS with a set of inventory data from several legacy information manufacturing systems. The overall approach shall be flexible and generic enough for application to other information data products within the Logistics and other AF Operational Support domains.

U.S. AIR FORCE

Outline

- AF Operations Support
- What Is Data Quality?
- Vision, Goals, Objectives, & Project Concept
- The Architecture of Data Quality
- DQ Process, Governance & Policy
- Enterprise Data Quality Management Service (EDQMS)
- Summary



U.S. AIR FORCE

What is Data Quality?

- **Quality** is frequently defined as:

  "Fit for purpose"

- Thus, good **quality data** can be defined as:

  "Data that is fit for its use"

- Good quality data exhibits these **characteristics**:
  - accurate, precise, complete, consistent, timely and authoritative

Integrity - Service - Excellence 5

U.S. AIR FORCE

What is Data Quality?

- **Accuracy**:
  - Correctness; Degree to which the reported information value is in conformance with the true or accepted value
- **Consistency/Validity:**
  - Degree of freedom from variation or contradiction
  - Degree of satisfaction of constraints (including syntax/format/semantics)
- **Completeness/Brevity:**
  - Degree to which values are present in the attributes that require them
  - Degree to which values not needed for decision making are excluded
- **Timeliness:**
  - Time/utility; Degree to which currentness of data values renders them useful
- **Pedigree/Lineage/Provenance:**
  - History of data origin and subsequent ownership and transformation
- **Precision/Certainty:**
  - Level of detail or exactness (vs. imprecise, approximate)
  - Confidence in value (vs. uncertain, probabilistic, or fuzzy)

Integrity - Service - Excellence 6

U.S. AIR FORCE

What is Data Quality?

- Good quality data is needed for:
  - Good decision making
  - Efficient and effective transaction processing.
- Data of known quality can be treated appropriately by decision support tools and transaction processing systems

U.S. AIR FORCE

What is Data Quality?

- For example, if you know what the quality of the data is, you can take a number of different actions:
  1. Go ahead and use it, knowing how reliable it is, and factoring that in
  2. Look for other corroborating or alternative sources for the data
  3. Clean the data up and use it
  4. Go back and fix the data operation(s) or producing system(s) to regenerate the data correctly
  5. Go back to the producing system(s) and improve them to prevent this type of problem in the future

U.S. AIR FORCE

To-Be Future State Vision

Enterprise Data Quality Management Strategy

- An Operational Support computing environment in which:
  - The quality of all data is defined and well known
  - Data exchanged between information systems is continuously monitored for quality
  - Data quality meta data is used to:
    - effectively manage ongoing system operations
    - support the clean up of problematic data for consumers (people and systems)
    - continuously improve overall information processing
    - better inform data owners, stewards and consumers to improve decision making at all levels

Integrity - Service - Excellence 9

U.S. AIR FORCE

DQ Project Objective & Goals

Objective:

Set up an Enterprise Data Quality Management Service (EDQMS), and then exercise it with a set of data from several legacy information manufacturing systems.

Goals:

- Provide an initial capability to manage the quality of data in the inventory area of the Logistics domain
- Leverage existing investments in Data Quality research and deployed infrastructure
- Approach shall be flexible and generic enough for application to other information data products within any domain and using any vendor products or legacy tools.

Integrity - Service - Excellence 10

U.S. AIR FORCE

AFIDQM Project Concept

AFIDQM

3. DQ Assessments for Inventory Data

1. Enterprise Data Quality Management Service (EDQMS)

2. DQ Infrastructure (GCSS-AF Data Services)

Exercise EDQMS for several DQ problems with sets of data from several legacy information manufacturing systems

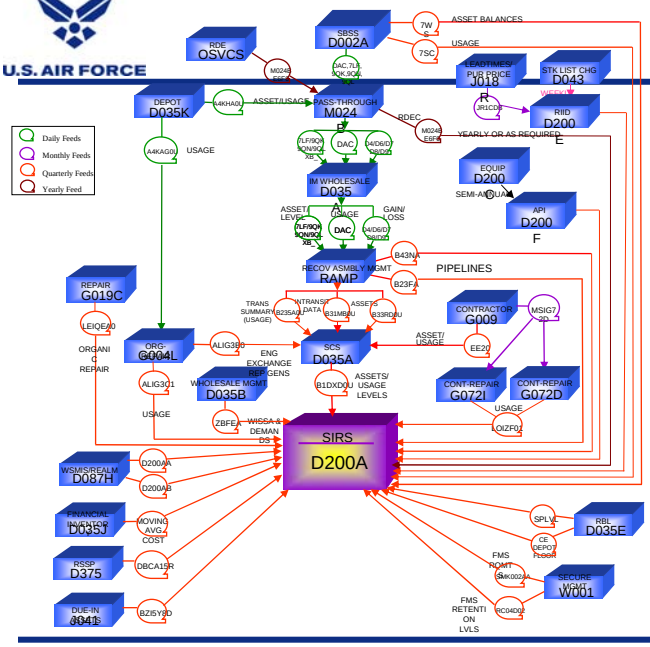
Establish end-to-end processes & governance constructs (EDQMS) that provide full **lifecycle management** of data quality

Upgrade of existing infrastructure to fully support EDQMS, as well as the loads and volumes of DQ metadata that a broad implementation of EDQMS will place on the infrastructure

Integrity - Service - Excellence

11

U.S. AIR FORCE

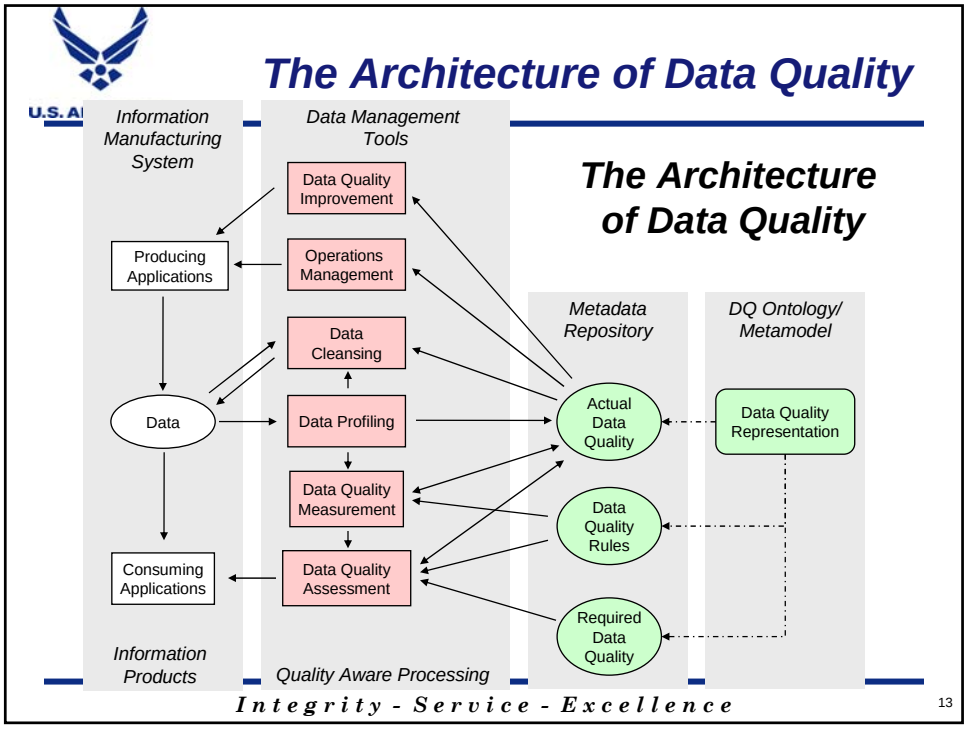


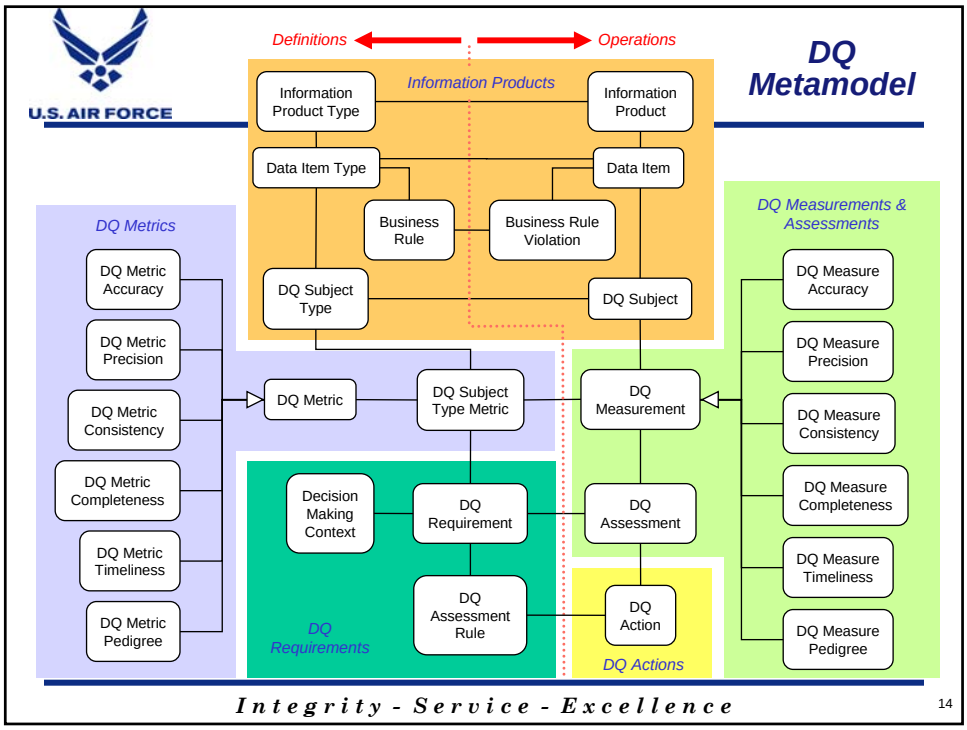
Example

Information Manufacturing Systems and Information Products for Spare Parts Forecasting

Integrity - Service - Excellence

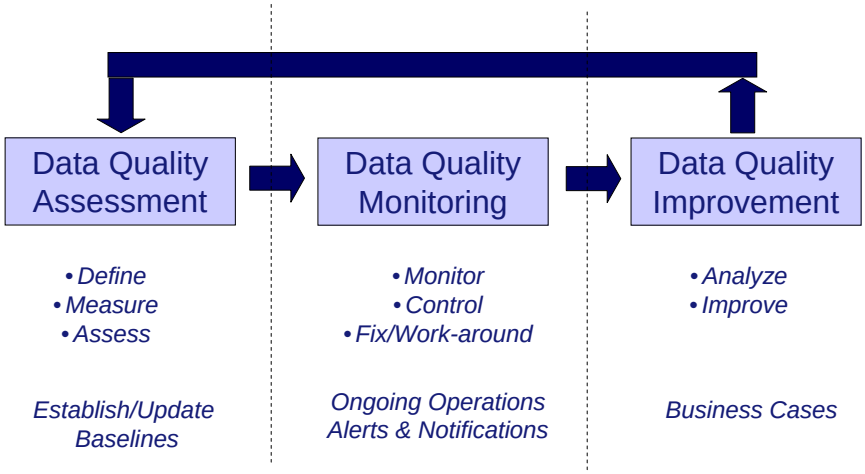
12


The Architecture of Data Quality
Information Manufacturing System
Data Management Tools
The Architecture of Data Quality
Data Quality Improvement
Producing Applications
Operations Management
Data Cleansing
Metadata Repository
DQ Ontology/ Metamodel
Data
Data Profiling
Actual Data Quality
Data Quality Representation
Data Quality Measurement
Data Quality Rules
Consuming Applications
Data Quality Assessment
Required Data Quality
Information Products
Quality Aware Processing
Integrity - Service - Excellence
13


Definitions
Operations
DQ Metamodel
U.S. AIR FORCE
Information Products
Information Product Type
Information Product
Data Item Type
Data Item
Business Rule
Business Rule Violation
DQ Subject Type
DQ Subject
DQ Metrics
DQ Measurements & Assessments
DQ Metric Accuracy
DQ Measure Accuracy
DQ Metric Precision
DQ Measure Precision
DQ Metric Consistency
DQ Metric
DQ Subject Type Metric
DQ Measurement
DQ Measure Consistency
DQ Metric Completeness
DQ Measure Completeness
Decision Making Context
DQ Requirement
DQ Assessment
DQ Metric Timeliness
DQ Measure Timeliness
DQ Assessment Rule
DQ Action
DQ Metric Pedigree
DQ Measure Pedigree
DQ Requirements
DQ Actions
Integrity - Service - Excellence
14

U.S. AIR FORCE

High Level DQ Business Process



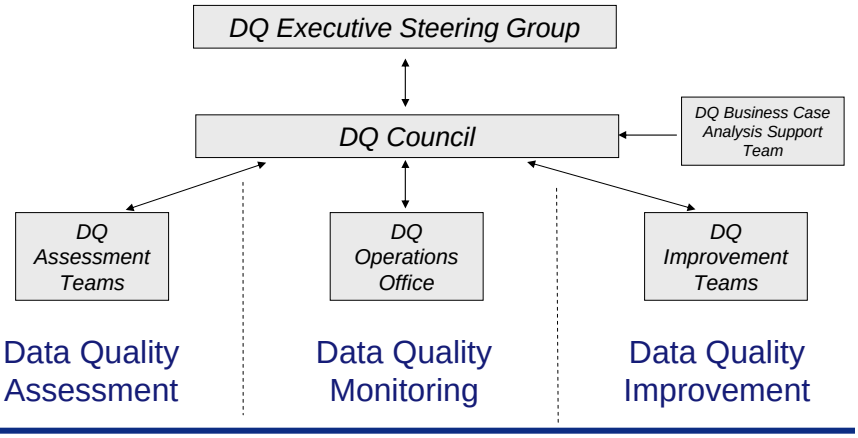
Integrity - Service - Excellence

15

U.S. AIR FORCE

High Level DQ Governance

The DQ governance construct should map closely to the process.



Integrity - Service - Excellence

16

U.S. AIR FORCE

Data Quality Policy

Reference: "Journey to Data Quality", Lee, Pipino, Funk & Wang, 2006, MIT Press.

- **Policy – A clearly articulated statement of vision and guidance for a viable, sustainable and effective data quality practice**
- **Disseminated/promulgated throughout the organization**
  - **Must be in place for the organization to remain engaged and to succeed in maintaining a viable, continuing data quality effort, which in turn proactively supports the organizations mission activities.**
  - **Ensures that efforts to attain and maintain high quality data and information are institutionalized, and not isolated to individual champions or departments.**
- **Addresses data quality practice, management, implementation, operations, metrics and standards, all at different levels of detail**
- **Will lead to continual improvement of the overall quality for use**

Integrity - Service - Excellence 17

U.S. AIR FORCE

EDQMS Process Phases

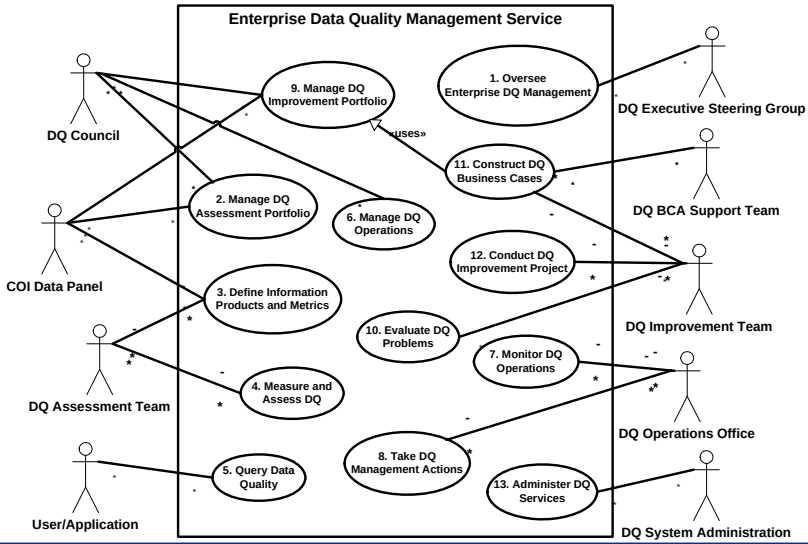
Cross Process Phases – There are currently five (5) cross process phases that have been identified for the EDQMS enterprise architecture.

- **MANAGEMENT & OVERSIGHT: oversees program initiatives, sets policies and procedures, secures funding for improvement projects, delineates data accountability, coordinates across Air Force enterprise, and oversees Stewardship & Coordination as well as Auditing and Compliance**
- **STEWARDSHIP & COORDINATION: establishes data standards, constructs enterprise vocabulary, determines data quality metrics and thresholds; oversees Data Quality (DQ) Operations & Improvement**
- **DATA QUALITY OPERATIONS: performs the actual data quality operations related to assessments, measurements, and monitoring performed by the organization utilizing various DQ tools**
- **DATA QUALITY IMPROVEMENT: improves data quality through analyzing sources of problems, cleansing data quality subjects, and correcting or re-engineering information manufacturing systems**
- **AUDITING & COMPLIANCE: measures and assesses compliance, as well as trust and confidence placed on EDQMS by it customers, by performing audits on EDQMS and the Data Quality organization**

Integrity - Service - Excellence 18

U.S. AIR FORCE

EDQMS Use Cases



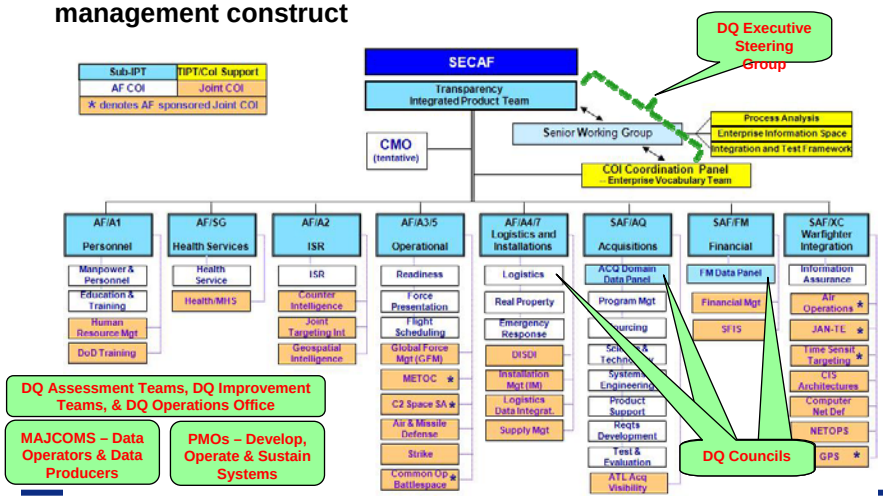
Integrity - Service - Excellence

19

U.S. AIR FORCE

EDQMS Governance – AF Transparency IPT (TIPT)

DQ governance construct should be aligned with existing AF data management construct

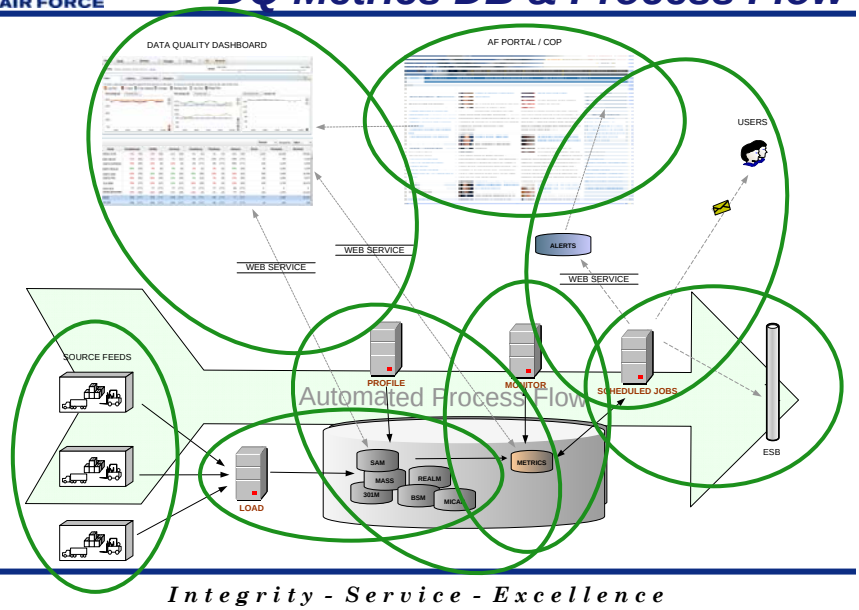


Integrity - Service - Excellence

20

U.S. AIR FORCE

GCSS-AF DS DQ Infrastructure DQ Metrics DB & Process Flow

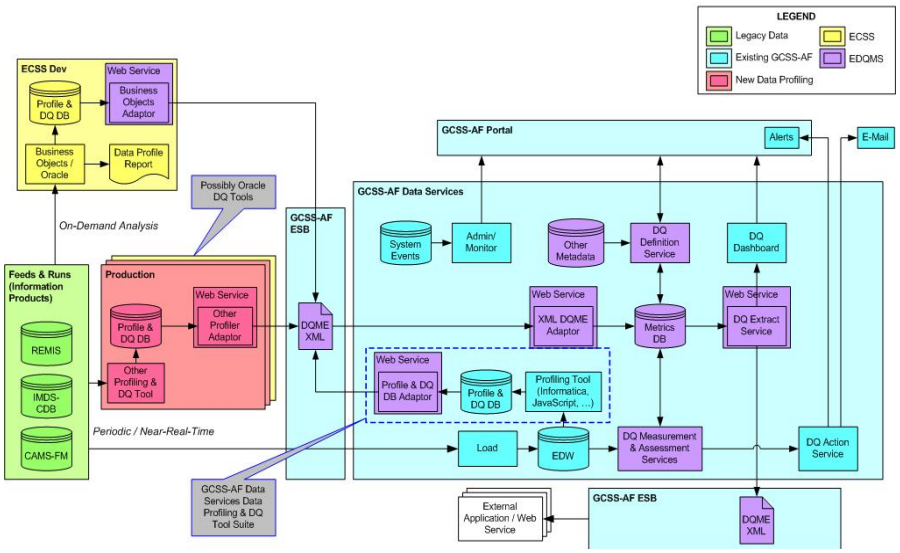


Integrity - Service - Excellence

21

U.S. AIR FORCE

EDQMS Systems Diagram

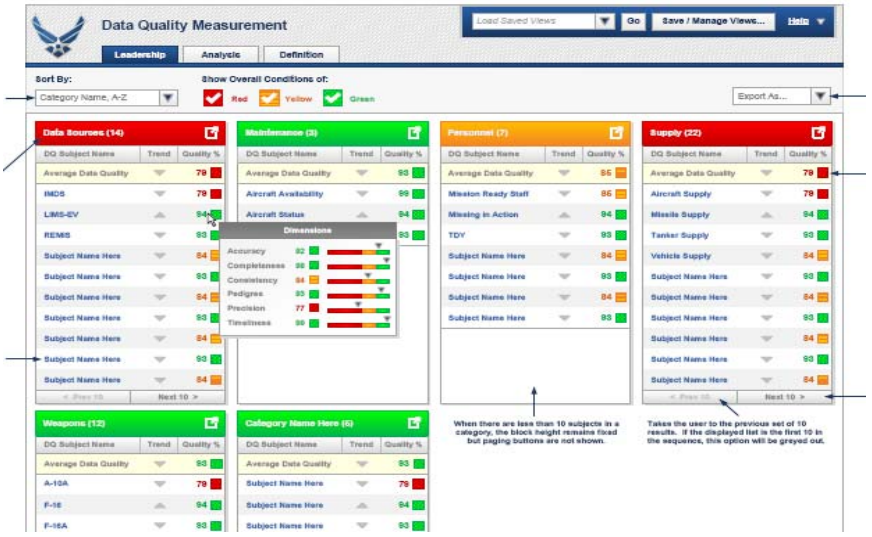


Integrity - Service - Excellence

22

U.S. AIR FORCE

EDQMS Dashboard Views

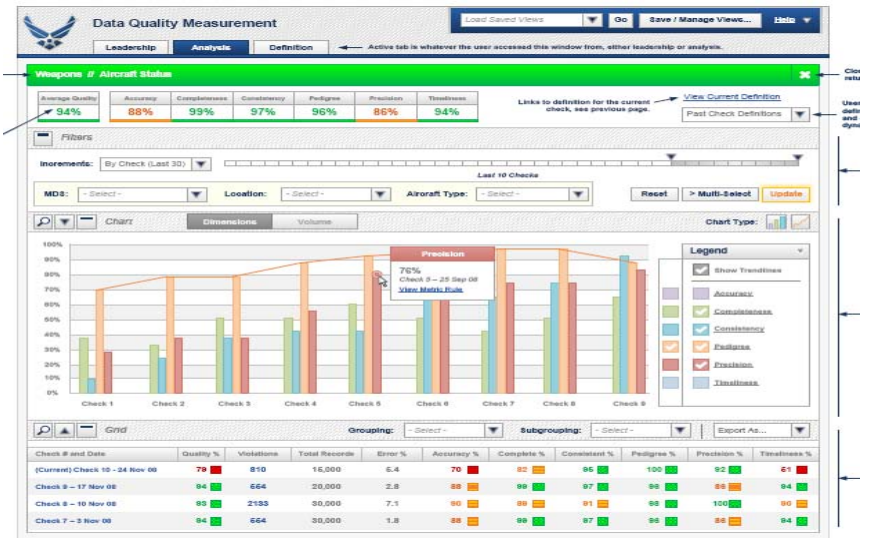


Integrity - Service - Excellence

23

U.S. AIR FORCE

EDQMS Grids, Charts & Graphs

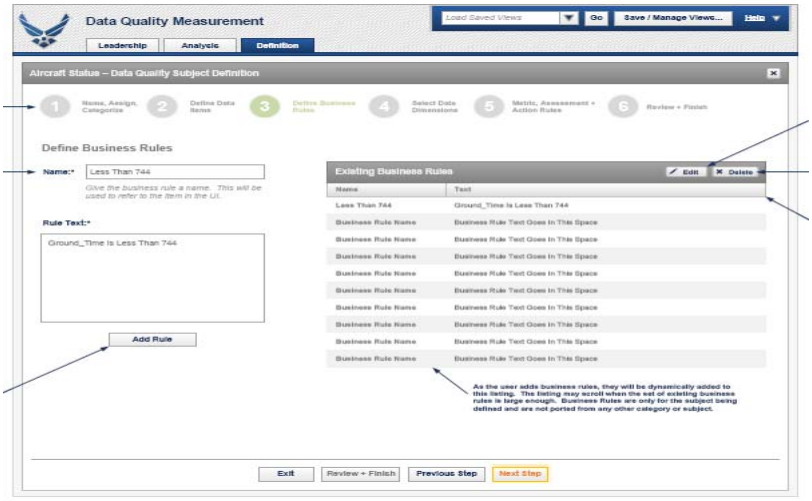


Integrity - Service - Excellence

24

U.S. AIR FORCE

EDQMS Definition Tools



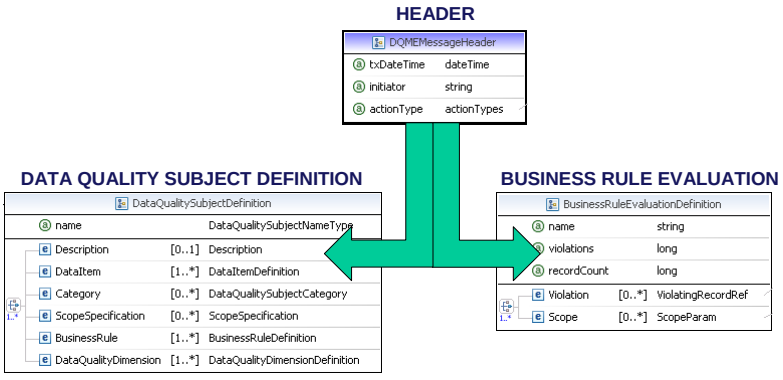
Integrity - Service - Excellence

25

U.S. AIR FORCE

DQME – Data Quality Metadata Exchange

The Data Quality Meta-Data Exchange (DQME) language is the communication format between the EDQMS infrastructure and all external systems, including Data Profiler Adapter implementations and other interested systems

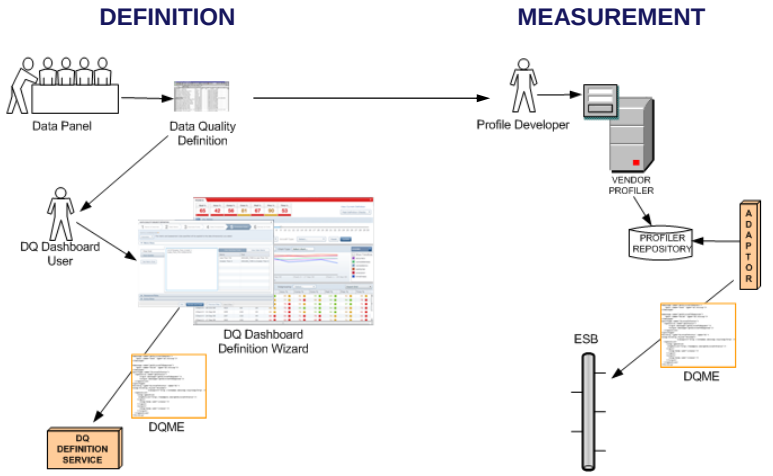


Integrity - Service - Excellence

26

U.S. AIR FORCE

DQME Process Flows

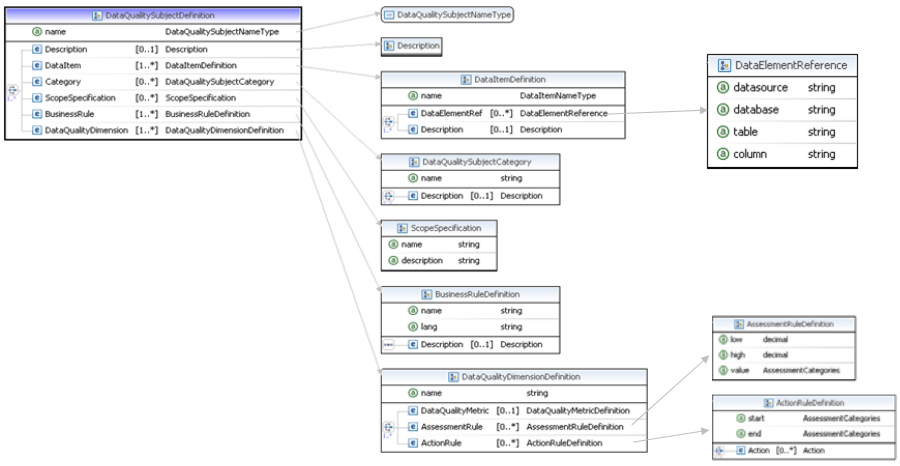


Integrity - Service - Excellence

27

U.S. AIR FORCE

Data Quality Subject Definition



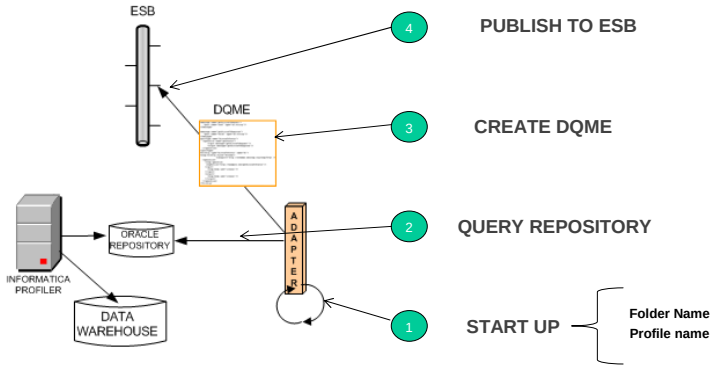
Integrity - Service - Excellence

28

U.S. AIR FORCE

DQME Adaptor

The DQME Adapter is a de-coupled component that collects business rule violations from a vendor-specific profiling or measurement tool and publishes the data in DQME format to the ESB.

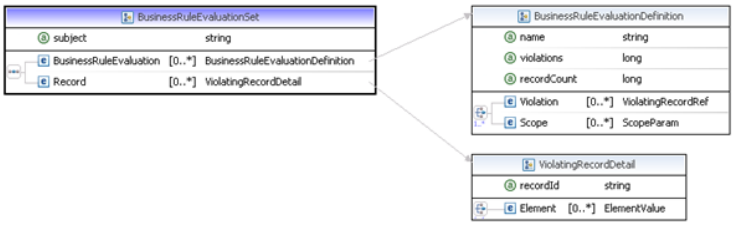


Integrity - Service - Excellence

29

U.S. AIR FORCE

Business Rule Evaluation



Integrity - Service - Excellence

30

U.S. AIR FORCE

Sample DQ Subjects

Master, Transactional, & Product Data

Master Data Objects			
1	Item	11	Customer
2	Organizations	12	Supplier
3	Work Center	13	Chart of Accounts
4	Vehicle	14	Person
5	Locator	15	Budget
6	Projects	16	Resources
7	Fixed Assets	17	Inventory
8	Equipment	18	Routes
9	Price List	19	Sourcing Rules
10	Carrier		

Transactional Data Objects			
1	Transportation Rates	10	Certifications
2	Resource Rates (Manufacturing)	11	Move Orders
3	Purchase Orders	12	Contracts
4	Shop Order	13	Invoice
5	Work Order	14	ECO
6	Requisitions	15	ECR
7	Sales Orders	16	Causal Factors
8	Inventory Balances	17	Forecast
9	Issue (ETAR)		

Product Data Objects			
1	BOM Structure	3	Serial Structure
2	Routings	4	Maintenance Program

Integrity - Service - Excellence

31

U.S. AIR FORCE

Summary

- We have a well defined service framework (EDQMS) for how DQ can be managed in an enterprise context
- It fits with our overall Enterprise Data Implementation Strategy
- It takes advantage of significant prior investment in DQ tools and capabilities
- It will be exercised with real data
- It supports a number of critically important current efforts, and is flexible enough to be expanded to many others

Integrity - Service - Excellence

32