Evolving Data Quality Management Capabilities in Complex Organizations: The Health Care Case

Bruce N. Davidson, Ph.D., M.P.H.

Director, Resource & Outcomes Management

Cedars-Sinai Health System

The Context
Cedars-Sinai Medical Center

- *Academic Medical Center/Health System*
- *Largest Non-Profit Hospital in the Western US*
- *950 Beds, 10,000 employees, 2000 MDs*
- *Basic Annual Statistics*
  - *55,000 inpatients*
  - *280,000 outpatients*
  - *55,000 ER visits*
  - *7,000 deliveries*

CEDARS-SINAI MEDICAL CENTER

Data Management Implications

- *Complex, information-intensive organization*
- *Distributed oversight responsibilities*
- *Transactional data systems populated as byproduct of patient care*
- *Information managed as departmental resource rather than as enterprise resource*

Evolving from this view…

MISSION

- Patient care
- Teaching
- Research
- Community Service

RESOURCES

- People
- Money
- Equipment

…to this view, over the last 10 years

MISSION

- Patient care
- Teaching
- Research
- Community Service

RESOURCES

- People
- Money
- Equipment
- **Information**

The "long and winding road" (1)

- *1997*
  - *DPG convened to address data crises*
- *1998*
  - *TDQM Summer Course*
  - *DQMWG Spun Off of DPG*
  - *IQ Survey, round 1*
- *1999*
  - *DQ Concept Kick-Off*
  - *IQ Survey, round 2*
  - *DQM Objectives first appear in Annual Plan*
- *2000*
  - *Big DQ Improvement Project*
  - *DQM Objectives appear again in Annual Plan*
  - *ROM Dept reorganization to capitalize on DQ framework*
- *2001*
  - *DPG & DQMWG Charters renewed*
  - *DQM Objectives again in Annual Plan*
  - *IQ Survey, round 3*

The "long and winding road" (2)

- *2002*
  - *DPG thinks about pro-active DQM infrastructure*
  - *ROM designated as data "clearinghouse" for approval of all clinical statistics reported out*
  - *JCAHO accreditation standards for MOI linked to DQM initiative*
- *2003*
  - *Data crisis: No P&Ls for 8 months*
  - *Initiate Data Warehouse Improvement Project*
- *2004*
  - *Propose DQMU (x2)*
  - *Implement Data Warehouse Improvement Project*
  - *DQMU funded as part of ROM*
  - *IQ Survey, round 4*
  - *Data crisis: No Management Reports for 5 months following new PM system implementation*
- *2005*
  - *DWIP Objectives appear in Annual Plan*
  - *DQMU staffed and DQM Program initiated*
  - *DWIP Project Plan institutionalized as primary focus for DPG*

The "long and winding road" (3)

- *2006*
  - *DQ Objectives appear in Annual Plan*
  - *DQ Objectives linked to executive management incentive compensation*
  - *Development of Data Certification Program for "High Priority" data elements*
  - *Pilot estimation of DQ ROI*
  - *Development of explicit criteria for resolving "High Priority" Data Quality Incidents*
- *2007*
  - *IQ Survey, round 5*
  - *117 Data Quality Incidents logged since initiation in February 2005, of which 82 have been resolved and closed*
  - *Collaboration with Internal Audit department relative to minimizing risk due to defective data*
  - *Continuing challenge to frame strategic issues in context of "managing information as an enterprise resource"*

What We Hope For

- *Quick results*
- *Unflagging support*
- *Universal cooperation*

What We Get

- *Often painfully slow progress with regular periods of stagnation, if not reversal*
- *Occasional bursts of support with frequent periods of inattention, or forgetfulness*
- *Gradually enhanced, but intermittent, cooperation*

What Does It Mean?

- ***Evolutionary process***
- ***Limitations of hierarchical control***
- ***Development of "shared mental models" = work***

Galaxy Data Quality Program
MIT IQ Industry Symposium
July 18-19, 2007

Ingenix
United Health Analytics
Galaxy – Shared Data Warehouse
Laura Sebastian-Coleman
IS Manager – Data Quality & End User Support

INGENIX®

Overview

- Ingenix and Galaxy
- Galaxy's DQ program
- Evolving business needs and the pace of change
- Data quality in relation to evolving business needs

INGENIX®

Ingenix Background

- A global healthcare information company
- Founded in 1996 to develop, acquire, and integrate some of the nation's best-in-class healthcare information capabilities
- Significant and rapidly evolving portfolio of tools and services now transform data into actionable, fact-based, technology-enabled decision support
- Ranked among the top 10 providers of informatics by *Healthcare Informatics* magazine in June 2006
- Today there is an Ingenix product at work in nearly every U.S. healthcare organization.
- Ingenix is a wholly owned subsidiary of UnitedHealth Group (UHG).

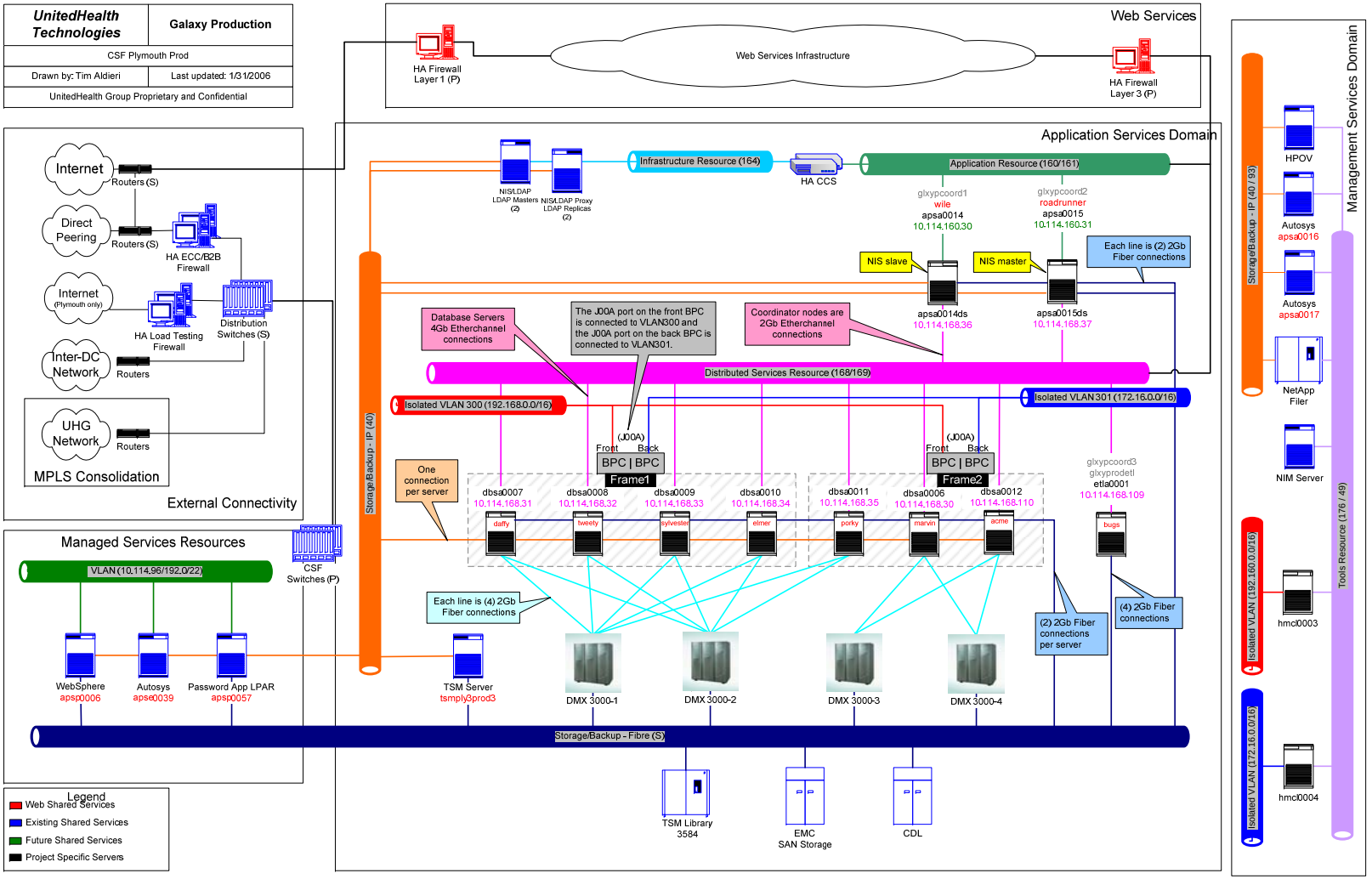
INGENIX®

Galaxy Overview

- Atomic Data Warehouse with transformations
- Integrates data from more than a dozen subject areas (claim, membership, customer, provider, etc.) across multiple sources
- Size
  - 350 source input files from more than 25 distinct internal and external sources (and counting)
  - 18 TB of data; 62 TB footprint
  - 3,159 attributes across 12,632 columns in 600 tables (and counting)
  - Largest table: more than 1.5 billion rows
    - 1,704,717,031 on Claim Statistical Service as of 5/3/07
- Usage
  - Over 1,000 registered users
  - 7,888 queries per day / 256,656 per month, on average
  - Ad hoc, scheduled queries, production extracts to applications and marts
  - Direct access to Galaxy via user-selected tools – Sagent is administratively supported

INGENIX®

Galaxy Physical Architecture

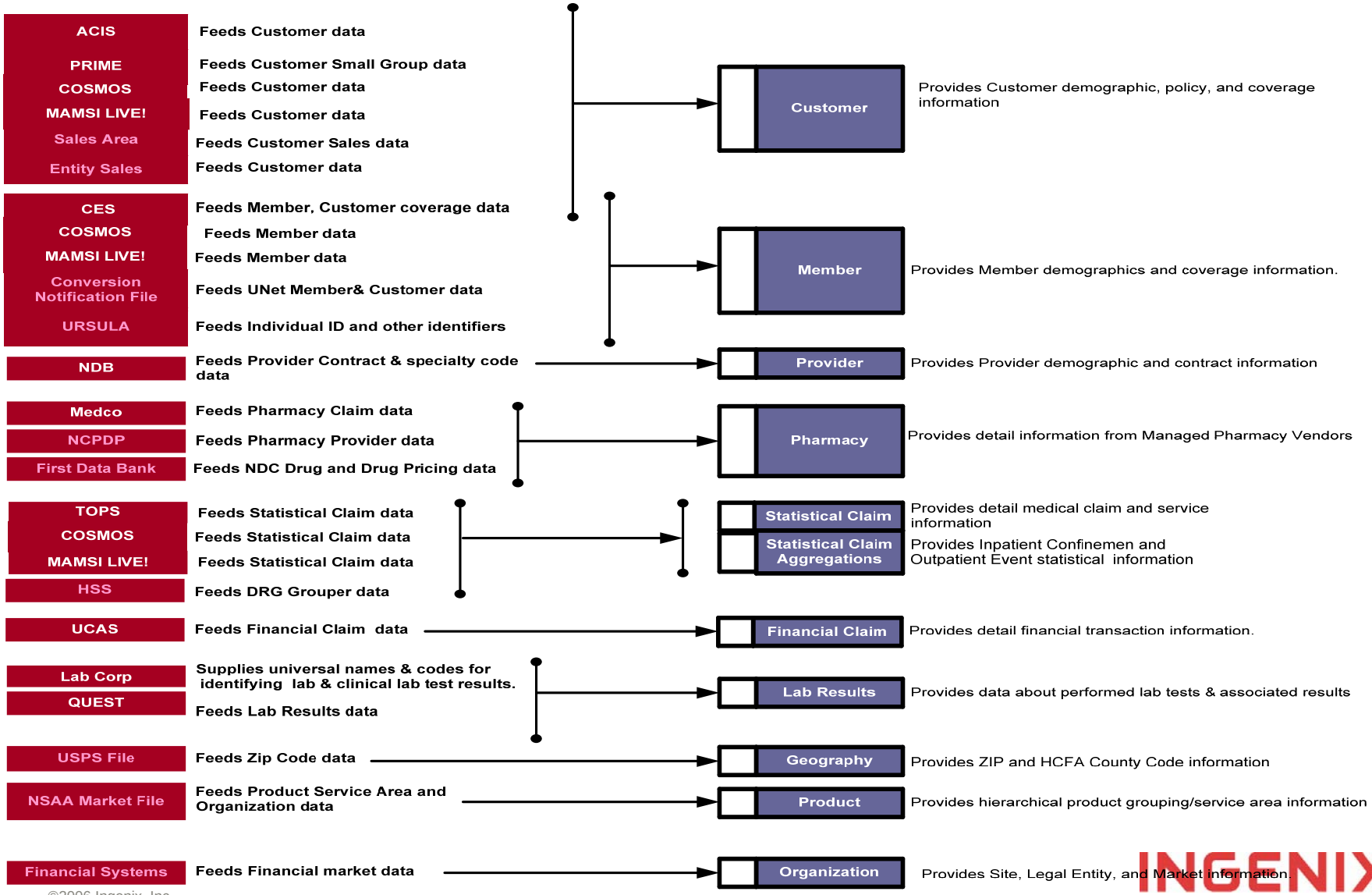


System Components

- Hardware
  - 7 IBM P-series Servers P575
  - 2 IBM P-series Servers P510
  - 1 IBM P-series Server P570
  - 4 EMC DMX 3000 Storage Cabinets
  - Additional supporting servers for Sagent, Autosys, etc.
- Software
  - UDB with DPF v8.2
  - AIX 5.3.0
  - DataStage/PX 7.0.1
  - Optiload 3.1
  - CoSort 7.5.3
  - Autosys 4.5
  - Sagent 4.5i

INGENIX®

Galaxy Source Systems & Subject Areas



Functions of Galaxy Data

Galaxy is the single source of truth for key business functions

- Medical Trend Analytics
- Pricing
- Provider Utilization & Profiling
- Appropriateness of Care
- Network Adequacy
- Care Management / Pattern of Care / Preventive Care
- Fraud & Abuse
- Customer Reporting
- HEDIS Reporting
- Member Demographics
- Product Penetration

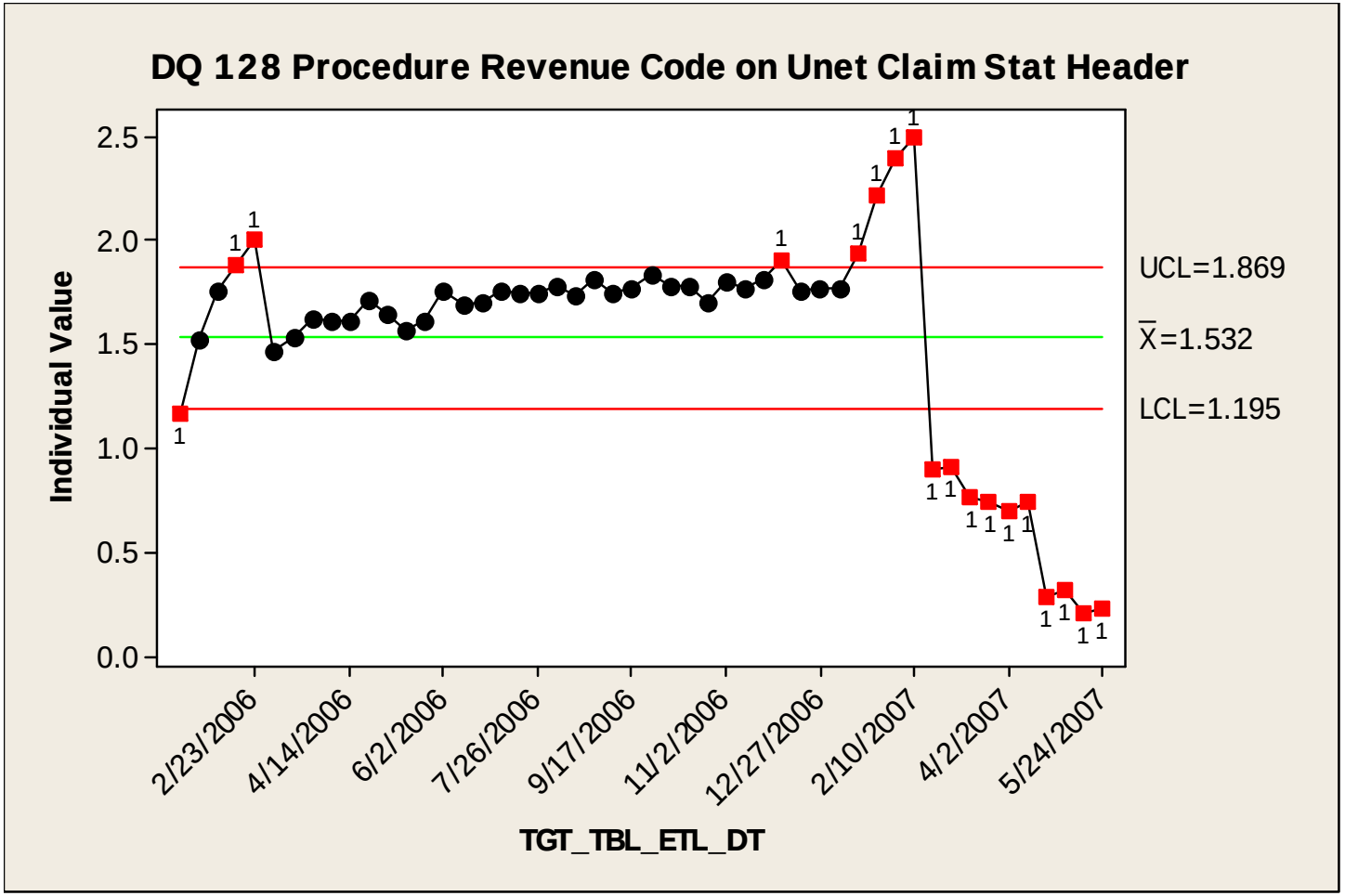
INGENIX®

Galaxy's Data Quality Program

- Management recognized need for DQ when Galaxy was launched
- Theoretical / methodological foundations
  - Correct data problems at the source
  - Data as a product
  - Statistical process control
- Primary functions of DQ program
  - Monitor, measure, and report on Galaxy's Data Quality
  - Recommend and implement actions based on findings
- Biggest initial challenge = establishing useful metrics
  - What to measure / how to measure
  - How to respond to the results of measurements
- 2003 Initiated metrics & reporting program
- 2004 Implemented first automated measures
- 2004-2007: Deliver weekly/cyclic, monthly, quarterly, semi- annual reporting through largely automated processes

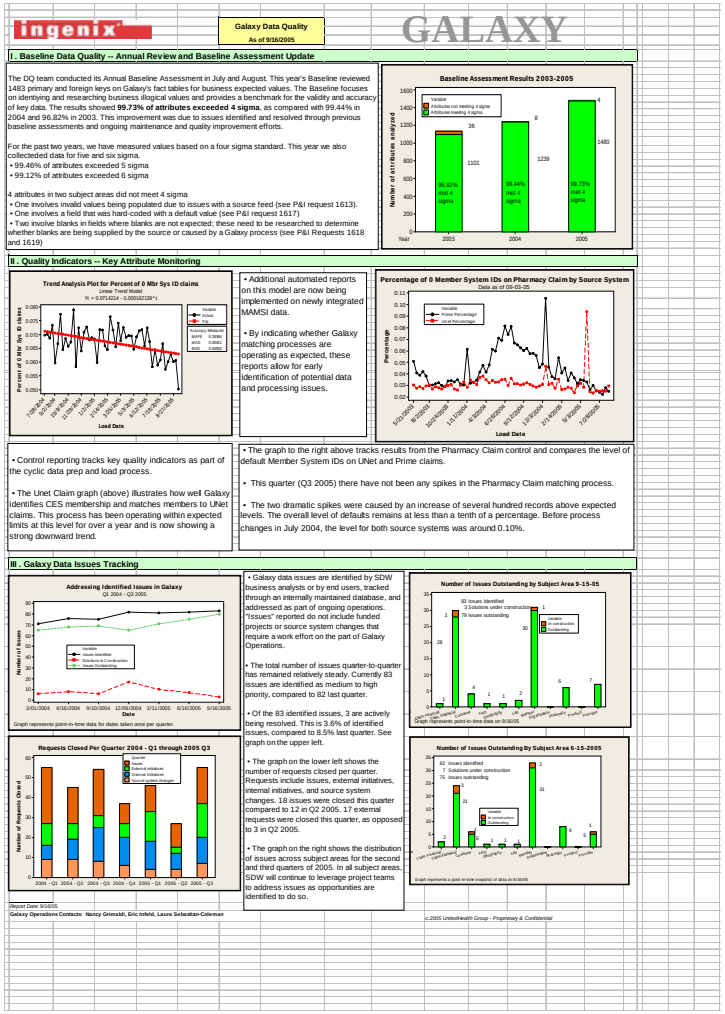
INGENIX®

Example of Weekly Measure



Quarterly Management Report

- Baseline Data Quality – annual review and baseline assessment
- Quality Indicators – key attribute monitoring
- Galaxy Data Issues Tracking



Current Situation

- Galaxy = a mature, enterprise data warehouse
- High demand for data and for organizational services
- Galaxy's DQ program also relatively mature
  - Defined metrics
  - Automated data collection
  - Regular reporting
  - DQ Community
- UHG growing, largely through acquisitions and partnerships
- Healthcare industry changing – relation of government to health care, new products, esp. consumer driven

INGENIX®

Pace of Change for Galaxy

- 2004
  - Galaxy integrated data from MAMSI, a United Health Group acquisition
    - Used the existing structure
    - 1+ year to integrate
- 2006
  - Integrated data from three new source systems
  - Developed a new subject area, Revenue
  - Significantly expanded Customer subject area
  - Responded to healthcare industry changes
    - Part D data
    - HRA (Health Reimbursement Account) data
- 2007
  - Integrate data from additional acquisitions
  - Expand the Revenue subject area
  - Continue to support the use and enhancement of existing data.
- 2008
  - Two major integrations already scheduled
  - Potential for several others

INGENIX®

Pace of Change for Galaxy DQ

- Biggest challenge
  - 2003 what to measure and how to measure
  - 2007 how to rapidly analyze and act on DQ data
- Baseline Assessment of Galaxy Data Quality
  - 2003
    - 800 person hours to pull and analyze data for first Baseline Assessment
    - Duration = more than 3 months
    - Measured 1137 attributes
  - 2006
    - Pulled 75% of data in less than 10 hours through an automated process
    - Measured 1506 attributes
    - Pull data quarterly
- Automated reports
  - 2004: 4 reports
  - 2007: 80 reports
  - Reports now implemented as part of standard development process.

INGENIX®

2007 – 2008 Key UHG Business Needs

- UHG acquisitions and partnerships –
  - More data for Galaxy
  - More users need access
- Users need data sooner –
  - Time to integrate data into Galaxy must be shortened
- Legacy data critical for ensuring reporting continuity and analytics –
  - Continued support is necessary
- Data consistency across sources critical for reporting continuity and analytics –
  - Integration methodologies need to promote and enforce consistency

INGENIX®

How to Respond?

- Data Quality included in set of changes to improve efficiency and agility
  - Common Interface – puts more responsibility on source systems for data quality
  - Gateway – changes how Galaxy prepares data.
- DQ measures
  - More comprehensive
  - Taken earlier in the process
  - More fully automated

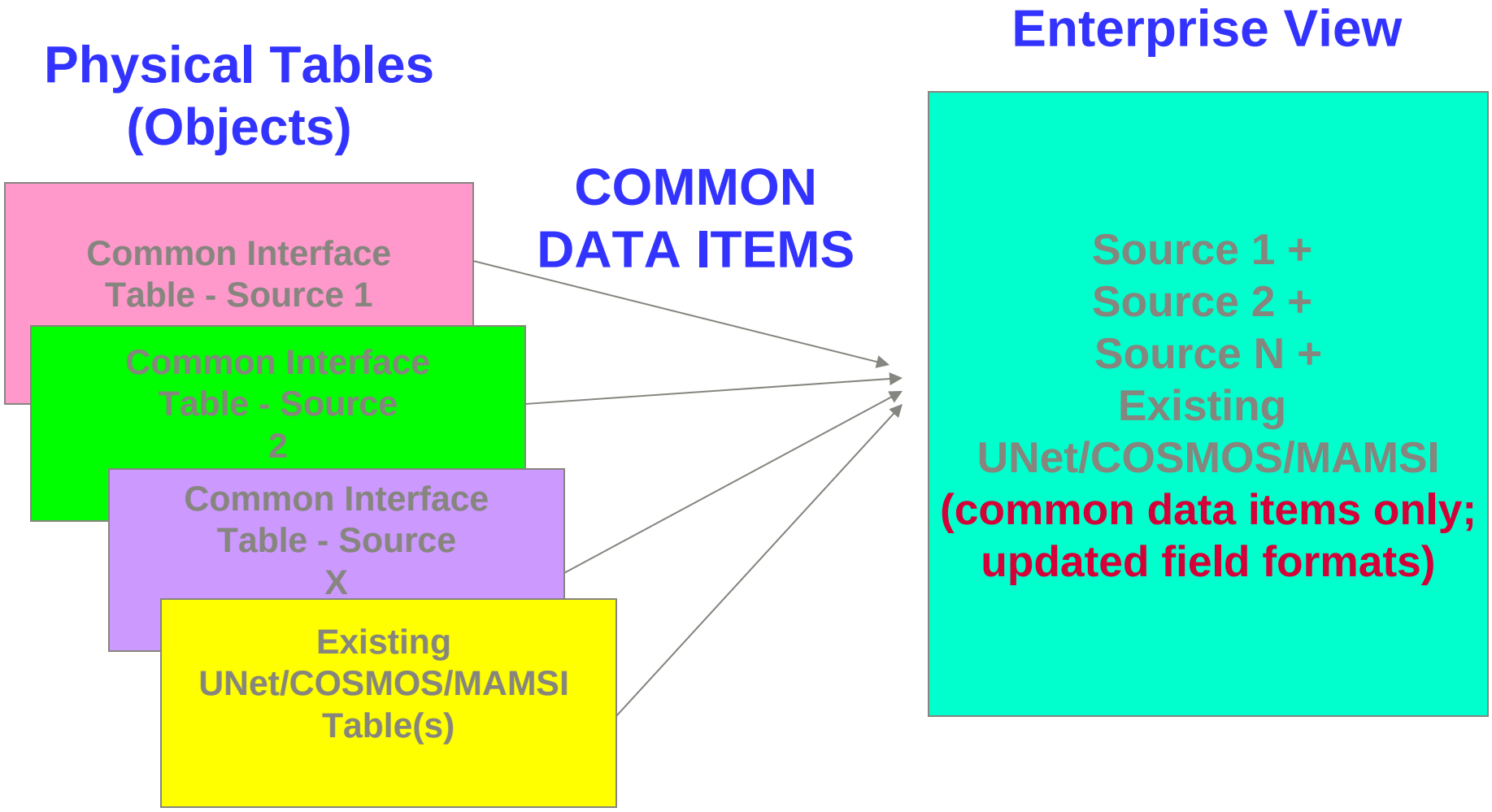
INGENIX®

Common Interface Approach

- Galaxy defines standard requirements and layouts for data
- Sources map to these requirements and feed to Galaxy
- Streamlined transformation/load into Galaxy
- Common model across the enterprise

INGENIX®

Common Interface Architecture – Views

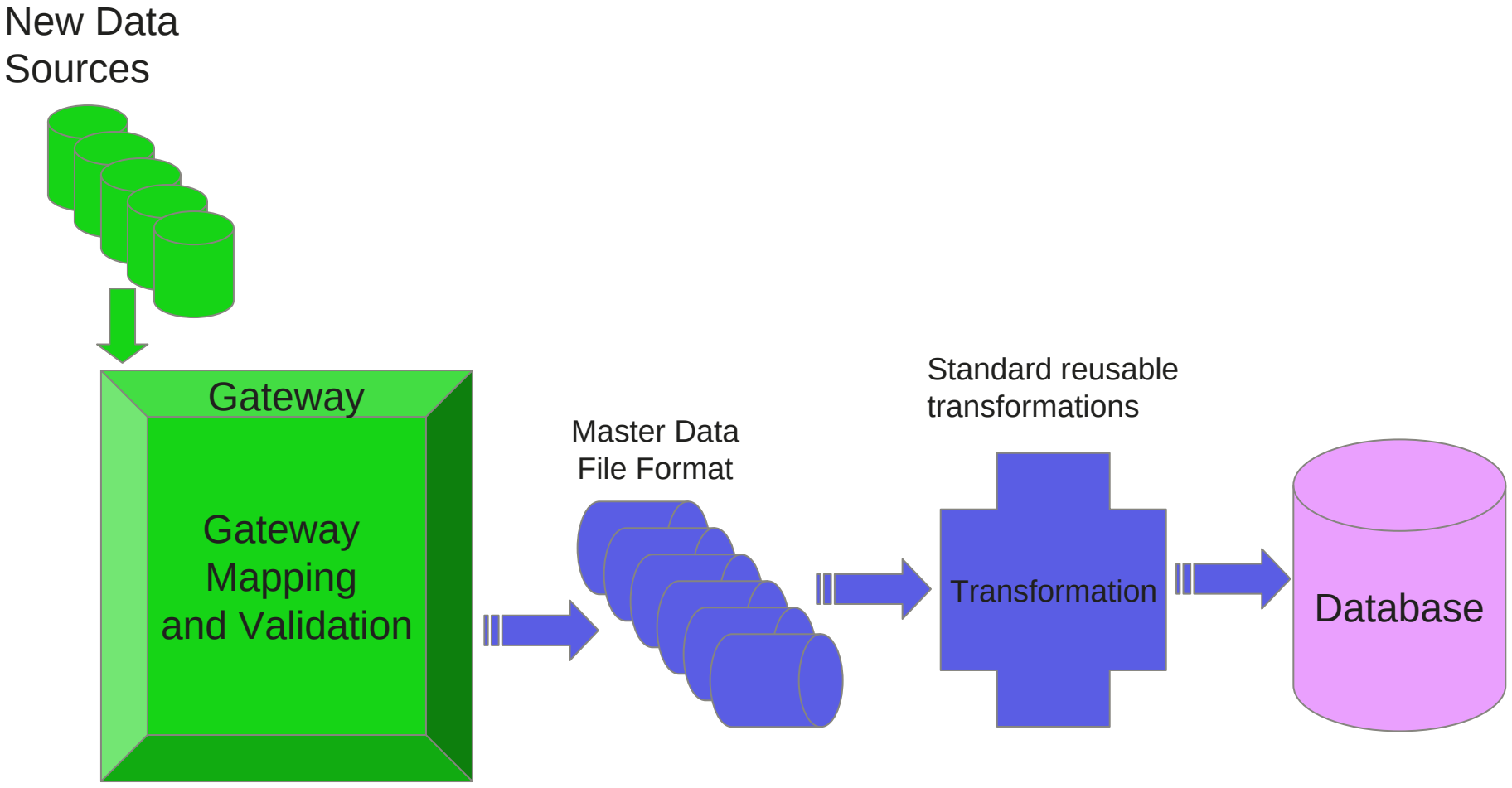


Gateway Integration Tool

- Facilitates mapping disparate data sources into a Master Data Definition
- Applies generic transformation logic to the output
- Utilizes reusable transforms
- Performs automatic code generation
- Ensures consistency across source-to-target mappings
- Provides true-to-code documentation
- Incorporates data quality modules
- Increases speed and reduces complexity of data integrations

INGENIX®

Gateway Integration Tool



INGENIX

Gateway – Data Quality Features

- DQ functions
  - Monitor and react to events in processing
  - Collect trend data
- Field validation
  - Data type checking
  - Value range checking
  - Valid value list checking
  - Assignment of default values
  - Informational, error and warning messages
- File validation
  - Format checking
  - Field counts / record length validation
  - Summary of field error and warning messages
  - Thresholds of summary counts of errors and warnings that allow job to be aborted if counts or percentages exceeded – generate alerts

INGENIX®

INGENIX

Back to Basic DQ

- Data in the warehouse is only as good as data in the source
  - Ensuring sources to supply better data through the Common Interface
- Manufacturing model: Data as a product produced through a process
  - Executing processes more consistently across the database through the Gateway
- Measure to improve
  - Gateway integrates and executes DQ measures consistently across the database.
  - Both tools measure ETL processes (timing of jobs, etc.) that affect other aspects of data quality from end-to-end

INGENIX®

DQ: Chicken or Egg?

- After 4 years – back to the beginning
  - Applying theory/methodology more fully
  - Applying at the beginning of integrations
  - Applying more comprehensively across the warehouse
- Major re-thinking of all Galaxy processes
  - Interacting with customers
  - Writing specifications
  - Obtain source files
  - Mapping source-to-target
  - Implementing ETL
  - Building physical tables
  - Taking DQ measures
- DQ still requires championing
- New problem: How to analyze and respond to findings from the data gathered through new process.

INGENIX®

INGENIX®

DSOWeb

Innovation in ETL

Transforming Raw Data into the Building Blocks of Intelligence

April 4, 2007

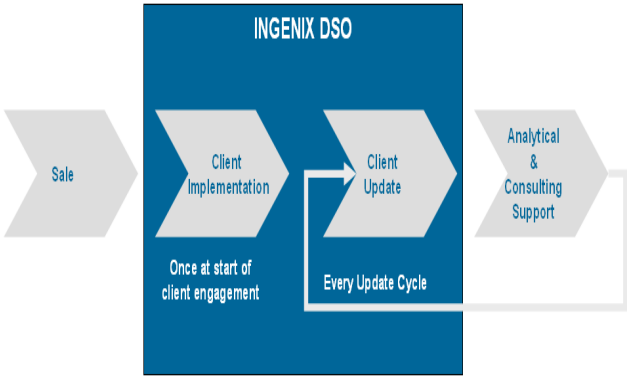
About Ingenix - Data Services Organization (DSO)

The Ingenix Data Services Organization (DSO) annually processes and integrates millions of service lines and other claim-related data from more than 500 diverse health and productivity related data sources for more than 125 large employers and hundreds of small employers. This translates to:

- 4250 feeds processed in 2005
- 8964 feeds processed in 2006
- 18,000+/- feeds processed in 2007

About Ingenix - DSO Processes

- Data is received across several media types related to eligibility, medical claim (includes vision and mental health), pharmacy claim, workers compensation, short-term disability, long-term disability, FMLA, lab results, disease management, health risk appraisal, payroll information, etc.
- Implementation is the design phase focusing on data layout, client account structure, initial data quality plan, conversion rules and successful completion of the first cycle.
- Update cycle is the periodic receipt of data from carriers/clients, focusing on data review, investigation/resolution of inconsistencies to ensure clean data, and loading onto analytical environment



Business and Operational Challenges

- Long turnaround time for data delivery
- Need for a cost effective data management solution
- Absence of streamlined data quality assessment and investigation process
- Absence of company-wide standardized data intake process
- Quality review process that is manual and subjective leading to errors, rework and decreased confidence in results
- High maintenance costs

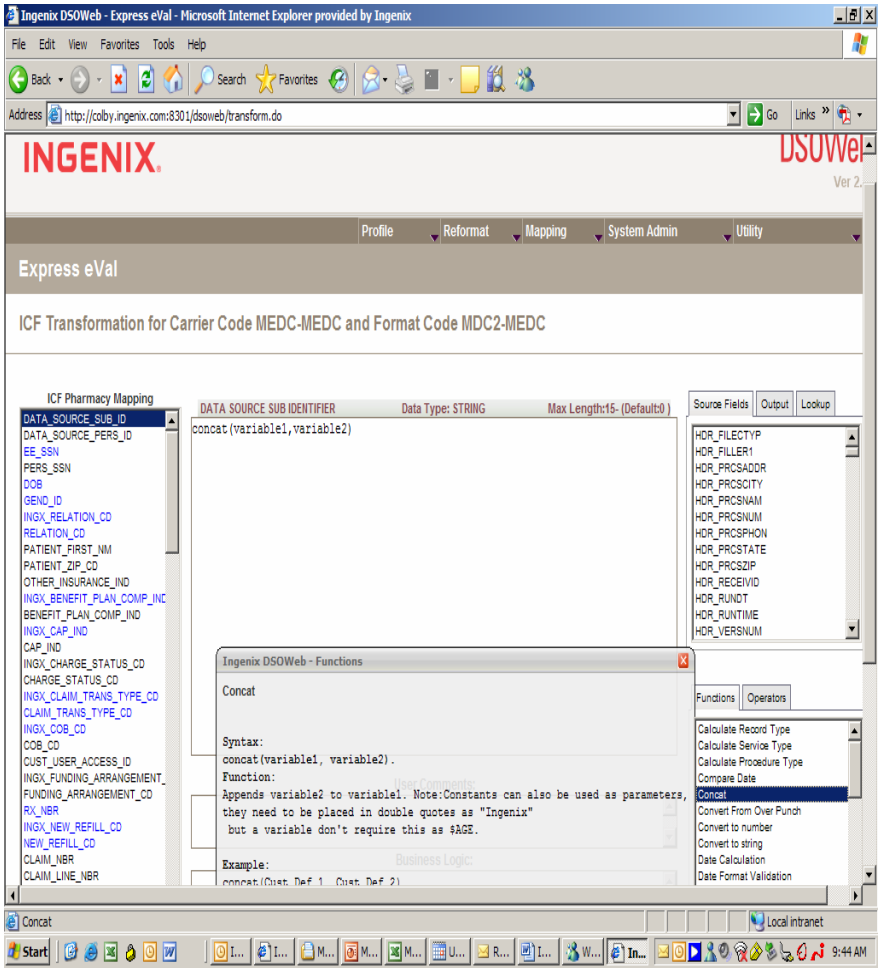
Business Needs

- Increase customer satisfaction through faster turnaround time and higher data quality
- Higher productivity through faster learning curve
- Elimination of manual intervention and continuous inspection of data
- Continually increasing automation
- Meet growing demand for faster data delivery and higher level of data quality
- Efficient data profiling and data management
- Flawless data delivery to analytical environments
- Ability to leverage transformed data across multiple Ingenix products

DSOWeb – Innovative ETL & DQ Solution

- Standardized data processing by data types – eligibility, medical, drug, disability, WC, FMLA, HRA, lab and disease management
- Standardized incoming data into common format by data type
- Automated data quality checks
- Automated data trending
- Automated file processing and job monitoring
- User 'friendly' interface
- Data quality and trending failure analyses
- Transformation and mapping functionalities
- Integrated data quality investigation functionalities
- Validated client and carrier details from profiles such as:
  - Employee status, employee type and types of coverage
  - Number of covered lives and products
- Operational and data quality metrics

DSOWeb - Express eVal – UI Driven Mapping Functionality



- Enables users to create source to common data mapping
- Point and click functionality; functions, operators and attribute selection easily added to transformation box
- Dynamic Help; displays syntax and example for each function
- On-screen listing of source attributes, mapped attributes and look-ups for transformations
- Text box to capture business logic for each transformation
- All transformations are stored in meta data tables and available for hard copy documentation and searches

DSOWeb – Data Quality

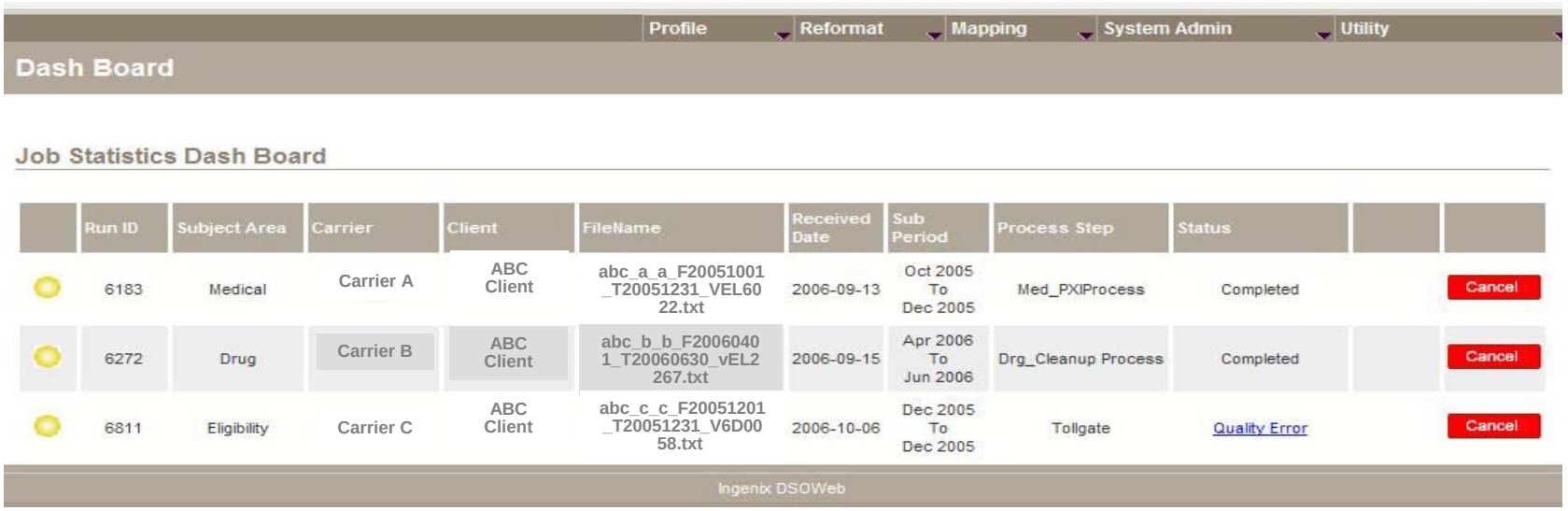
- Data Quality Rule engine systematically compares data results to rule thresholds and only flags exceptions for DSOWeb user intervention
- Metadata Driven Rules Engine - Quick turnaround for adding, deleting or modifying rules
- More than 7,000 data quality rules tailored to each data type
- Different rule types for column property enforcement, structure enforcement and business rules enforcement
- Formulated from verified client claims, eligibility, lab results, HRA and workforce productivity data experience
- Based on record distribution and benefits paid threshold

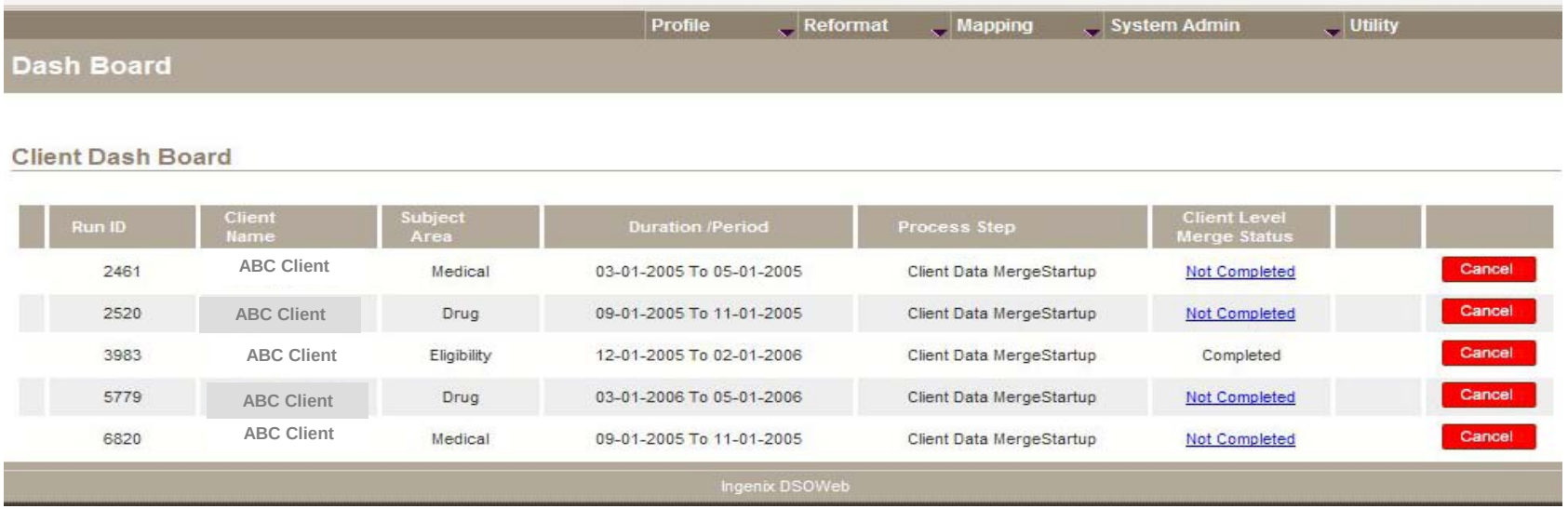
DSOWeb – Trending

- Flags unexpected trends in data. Two levels are:
  - Month-to-date trending with the previous month's data
  - Year-to-date trending with the previous year's data.
    - For year-to-date, the trending is done up to the month the data is being processed
    - **Example**: If the plan begins in January and the current processing month is June, then trending is performed between received year's January-June and data for January-June of the previous year.
- Metadata driven trending rules engine; quick turn around for adding, deleting or modifying rules
- Configurable tolerance ranges for trending rules
- Examples of tolerable ranges

Record Count	Passing Range
Less Than 10000	5% to +5%
10001 -25000	-3% to +3%
Greater Than 25000	-1% to +1%

DSOWeb – Job and File Processing Dashboard





DSOWeb – UI Driven DQ Failure Analysis and Decision Making

Data Quality Overrides

RunId 6749 File Name Abc_car1_car2_F20051001_T20051231_VEL6447.txt

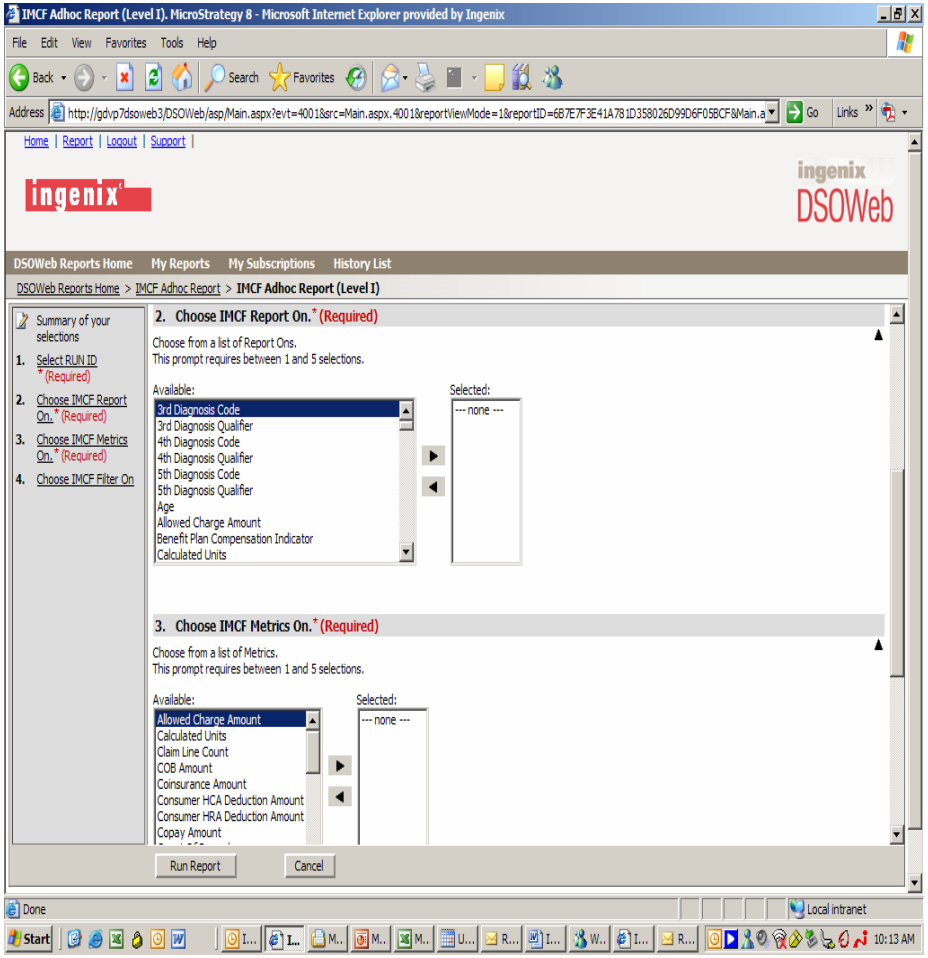
Client ABC Client Carrier E Car1

Recieved Date 2006-10-04 Sub Period Oct 2005 To Dec 2005

Rule Level: Claims Before Merge | Rule Type: 1 Way Frequency | Apply Filter | Back

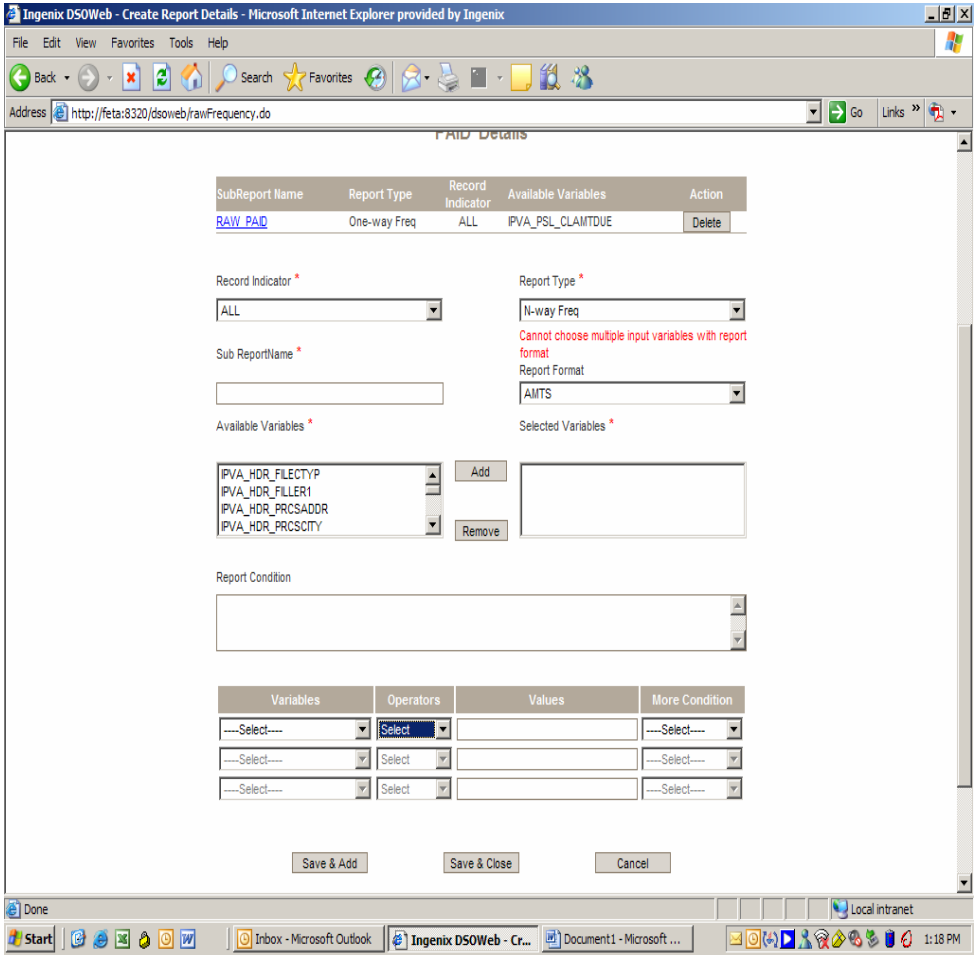
Rule Id	Override Type	Rule Type	Rule Level	ICF Variable(Condition)	Upper Limit	Lower Limit	Actual(%)	Actual Count	Override Level	Reason For Override
5025	Permanent	1 Way Frequency	Claims Before Merge	INGENIX RELATION CODE = 3	21.0	11.0	22.15	19913	Data Feed	OK for Client X - Just outside of norm
5037	Permanent	1 Way Frequency	Claims Before Merge	INGENIX CLAIM TRANSACTION TYPE CODE = 2	4.5	1.0	0.0	0	Carrier-Format	Carrier Y does not send adjustments, only
5041	Current Run	1 Way Frequency	Claims Before Merge	INGENIX COB CODE = 2	95.0	80.0	98.06	88171	-Select-	Slightly higher than norm. Will moniter
5046	Permanent	1 Way Frequency	Claims Before Merge	INGENIX DETAIL CATEGORY CODE = 10	0.75	0.01	0.0	2	Data Feed	OK - expected to have little or no experience
5046	-Select-	1 Way Frequency	Claims Before Merge	INGENIX DETAIL CATEGORY CODE = 45	0.75	0.01	4.44	3992	-Select-	
5046	-Select-	1 Way Frequency	Claims Before	INGENIX DETAIL CATEGORY CODE = 51	0.75	0.01	0.0	0	-Select-	

DSOWeb – Integrated DQ Investigation Tool



- Data investigations and validation environment integrated within application
- Uses BI tool allowing users to query converted data without needing to understand the data structure or a query language
- Ad-hoc and standard reports by data type
- Automatically schedules and creates standard reports with each production update

DSOWeb – Raw Data Analysis Tool



- Raw data investigation tool integrated within application
- Uses home grown tool that generates SAS queries without the need to understand SAS
- Ability to save queries for re-use
- Automatically emails formatted results to end user for printing or sharing

DSOWeb – Benefits

- Cost-effective data management solution
- Facilitates timely data investigation
- Faster turnaround time for data delivery
- Automated process eliminates human errors
- Streamlined process ensures scalability
- Metadata based DQ engine; ease of rules maintenance
- Human intervention only targeted at exception cases; increases productivity and maximizes quality output
- Timely issue resolution
- Metrics on data quality and operational processes can be reported as needed with current data
- Reduces dependency on technologists in transformation process

Ingenix - DSOWeb

- Thank You

The MIT 2007 Information Quality Industry Symposium

Cambridge, Massachusetts, USA

Proceedings of the MIT 2007 Information Quality Industry Symposium

9 - 10:30 AM **Session 1B: Federal Data Architecture** **E51-372**

Moderator: Skip Slone, Lockheed Martin

1. Adel Harris, Citizant, Inc
2. Mark Amspoker, Citizant, Inc
3. Burton Cutting, Royal Bank of Canada