



# **Proceedings of the Fourth MIT Information Quality Industry Symposium**

Massachusetts Institute of Technology  
Cambridge, MA, USA

July 14-16, 2010

Edited by

**WooYoung Chung**  
MIT Information Quality Program

Copyright © 2010 MIT Information Quality Program, Massachusetts Institute of Technology

All presentations included in these proceedings are published with explicit permission from their author or copyrights holder. These presentations are protected by the international and U.S. copyright law. Any copying, distribution, exhibition in public or private meetings, and/or pursuing any derivative work, by any third party, in part or as a whole, is strictly prohibited without explicit written permission of the original copyrights holder. Any third party who wishes to quote or refer to the information presented in these proceedings must comply with the fair use provisions of the U.S. copyright law 17 U.S.C. §107 and must properly cite the author and these proceedings as its source.

## FOREWORD

**WELCOME** to the *4<sup>th</sup> MIT Information Quality Industry Symposium (MIT 2010 IQIS)*. The last ten years have proven exciting for data quality practitioners, as more and more organizations have come to understand the critical importance of high-quality data. Data quality efforts in large enterprises often fail to deliver on the expectations of the business. As with any service provision, the key to avoiding such disappointment is to look at the problem from the beginning through the eyes of the business.

Many have made vast improvements serving as role models for others. The next ten years promise to be even more exciting. Leading organizations including the MIT Information Quality community are coming to realize that data, just like capital and people, are assets that must be explicitly managed, as any valuable corporate asset would be.

In order to produce an event that meets the needs of data and information quality practitioners, the MIT IQIS organizing committee has researched extensively with a range of data experts in data management, data warehousing, business intelligence, customer information and information management, across a range of industries and sectors.

We are honored and privileged to have Mr. Michael Krieger, Deputy Chief Information Officer, Office of the Army CIO/G6, to give the keynote address at the Symposium opening, as well as Mr. John Bottega, Vice President and Chief Data Officer of the Federal Reserve Bank of New York to keynote on Thursday.

This event is co-sponsored by the MIT Information Quality Program and the U.S. Army, and supported, in part, by Burton Group, Citizant, The Data Governance Institute, Harte-Hanks Trillium Software, Symbolic Systems, and Talend.

We would like to express our gratitude to Forea Wang, Allen Moulton, Rith Peou, John Garner and other MIT staff for their assistance in producing the Symposium proceedings, and managing the various aspects of the Symposium operations. Thanks are also due to the University of Arkansas at Little Rock for managing the Symposium website, in particular Kasia Dziewanska. The final Symposium program and other information regarding the Symposium are available at <http://mitiq.mit.edu/IQIS>.

**MIT 2010 Information Quality Industry Symposium**  
**Organizing Committee**

**General Co-Chairs**

Suzanne Acar	Federal Bureau of Investigation
Peter Anlyan	Anlyan Consulting, LLC
Linda Kresl	MIT Information Quality Program

**Program Co-Chairs**

Joseph Bugajski	Burton Group
Garry Darrer	Getinge USA, Inc.
Willa Pickering	Lockheed Martin Corporation
Charles McKinney	Freddie Mac

**Proceedings Chair**

WooYoung Chung	MIT Information Quality Program
----------------	---------------------------------

**Symposium Manager**

Carol Chapman	American National Insurance Company
---------------	-------------------------------------

**Publicity Co-Chairs**

Peter R. Benson	Electronic Commerce Code Management Association
Steve Sarsfield	Talend

**Advisor Committee**

Bruce Davidson	Cedars-Sinai Health System
Yang Lee	Northeastern University
Stuart Madnick	Massachusetts Institute of Technology
Richard Wang	Office of the Army CIO/G6 & MIT IQ Program

# MIT Information Quality Industry Symposium

Location

Massachusetts Institute of Technology  
Tang Building (E51) in the MIT East Campus  
70 Memorial Drive, Cambridge, MA, USA  
Dress Code: Business casual

Date

July 14, 2010 – July 16, 2010

Official Website

<http://mitiq.MIT.edu/IQIS/2010>  
Contact: Dr. Richard Wang  
Deputy Chief Data Officer  
Chief Data Quality Officer  
U.S. Army

Director, MIT Information Quality Program (on leave)

WEDNESDAY, July 14, 2010		
8:00-9:00 AM	CONTINENTAL BREAKFAST & CHECK-IN	Room
9:00-10:30 AM	<b>Session T1A</b> Tutorial Workshop: Business Intelligence  Textual Integration -- Some Business Examples ..... William H. Inmon Inmon Consulting Services	E51-345  pg. 1
	<b>Session T1B</b> Tutorial Workshop: Data Governance  Data Governance with a Focus on Information Quality Gwen Thomas President The Data Governance Institute	E51-361
	<b>Session T1C</b> Tutorial Workshop: ISO 8000 Standards  The NATO Codification System: Improving Data Quality through ISO Standards 22745 and 8000 ..... Steven E. Arnett Chief, U.S. National Codification Bureau Defense Logistics Agency	E51-372  pg. 11

	<p><b>Session T1D</b> Tutorial Workshop: Data Quality Solution</p> <p>1. Operational Data Quality Applied Michele Goetz Director Product Marketing Trillium Software</p> <p>2. Affordable Data Management ..... pg. 50 Sebastiao Correia R&amp;D Engineer, Talend, France Steve Sarsfield Product Manager, Talend</p>	E51-376
10:30-11:00 AM	Coffee Break	
11 AM-12:30 PM	<p><b>Session T2A</b> Tutorial Workshop: IQ and Entities</p> <p>Entity &amp; Identity Resolution ..... pg. 68 John R. Talburt Director of ERIQ Laboratory University of Arkansas at Little Rock</p>	E51-345
	<p><b>Session T2B</b> Tutorial Workshop: Data Archiving</p> <p>Data Quality and Database Archiving: The Intersection of Two Important Data Management Functions ..... pg. 101 Jack Olson Chief Executive Officer SvalTech</p>	E51-361
	<p><b>Session T2C</b> Tutorial Workshop: Enterprise Architecture</p> <p>Enterprise Architecture's Identity Crisis: Caution Sign for the EA/IQ Relationship ..... pg. 120 P. Kathie Sowell President Custom Enterprise Solutions, LLC</p>	E51-372
	<p><b>Session T2D</b> Tutorial Workshop: Business Quality</p> <p>Quality Principles, Processes and Techniques Applied to Information Quality Management ..... pg.132 Larry P. English President Information Impact International, Inc.</p>	E51-376

12:30-2:00 PM	<p>LUNCH &amp; KEYNOTE SPEECH</p> <p><b>Chairs' Welcome and Opening Remarks</b></p> <p>Symposium Co-Chairs Suzanne Acar, Peter Anlyan, Linda Kresl</p> <p>Program Co-Chairs Willa Pickering, Joseph M. Bugajski, Garry Darrer, Charles McKinney</p> <p>UALR MSIQ/PhD Update John Talburt &amp; Rolf Wigand</p> <p>ACM-JDIQ update Yang Lee</p> <p>IQ Society Update QIBRAS (Brazil), ExIQ (France), DGIQ (Germany), ArgIQ (Argentina), IAIDQ</p> <p>MIT 2011 IQIS Announcement Joseph Bugajski (Business &amp; Finance) Peter Anlyan (Health Care) Suzanne Acar (Government)</p> <p><b>Keynote Speech</b></p> <p><i>Data -- A Force Multiplier for Effective Decision Making</i> ..... pg. 163</p> <p><b>Mr. Michael Krieger</b> Deputy CIO, Office of the Army CIO/G6</p>	E51-345
2:00-2:30 PM	COFFEE BREAK	
2:30-4:00 PM	<p><b>Session 1A</b> Government IQ and Metrics for Shared Information Session Chair: Willa Pickering, Lockheed Martin Corporation</p> <p>1. Paradox in Network Data Quality of Service: Consumer Perspective vs. Provider Perspective ..... pg. 164 Susmit H. Patel Senior Principal Engineer The MITRE Corporation</p> <p>2. Flexible and Generic Data Quality Metrics ..... pg. 174 David Becker Principal Information Systems Engineer The MITRE Corporation</p>	E51-345

	<p><b>Session 1B</b> IQ in Medical Informatics I  Session Chair: Peter Anlyan, Anlyan Consulting, LLC</p> <p>1. Data Quality and Stewardship in the Veterans Health Administration ..... pg. 175  Mark Love  Data Stewardship Program Manager  Veterans Health Administration (VHA)  Sara Temnitz  Business Product Manager  Office of Information, Health Data and Informatics,  Data Quality Program  Department of Veterans Affairs</p> <p>2. Aligning Multiple Data Sources for Quality, Regulatory Compliance,  and Documentation Improvement around Hospital-Acquired  Conditions ..... pg. 191  Brian Taylor, PhD  Director of Clinical Analytics  New York Presbyterian Healthcare System  Wazim Narain, MPH  Specialist, Data Analytics / Business Solutions Group  New York Presbyterian Healthcare System</p>	<p>E51-361</p>
	<p><b>Session 1C</b> Data Governance  Session Chair: Joseph M. Bugajski, Burton Group</p> <p>1. Effective Information Governance ..... pg. 204  Steve Adler  Director, IBM Data Governance Program  IBM</p> <p>2. Improving Your Business IQ and Governance ..... pg. 205  Linda Kresl  Data Warehouse Manager  Yahoo!</p>	<p>E51-372</p>
	<p><b>Session 1D</b> Financial Industry Enterprise Architecture  Session Chair: Charles McKinney, Freddie Mac</p> <p>1. Enterprise Architecture and Business Transformation: Engaging  the Executive Suite ..... pg. 217  J. McDuffie Brunson  Managing Director  KPMG</p> <p>2. Model Driven Message Interoperability ..... pg. 218  Mark Eisner  Chief Technology Officer  Firestar Software</p> <p>3. Financial Track Overview: IQ and Regulatory Reform ..... pg. 227  Charles McKinney  Senior Director, Advanced Analytics  Freddie Mac</p>	<p>E51-376</p>



4:00-4:30 PM	COFFEE BREAK	
4:30-6:00 PM	<p><b>Session 2A</b> IQ Practice in Government Agencies  Session Chair: Judith A. Pinsky, CECOM LCMC Software Engineering Center</p> <ol style="list-style-type: none"> <li>1. Implementing Data Quality: One Data Element at a Time  Frank Ponzio  President and CEO  Symbolic Systems, Inc</li> <li>2. Standard Ontologies for Improved Data Quality  Jim Schoening  Team Leader of the Net-Centric Data Transformation Team  Data Strategy Center of Excellence</li> <li>3. Putting IT All Together: Making Data Visible, Accessible, Trusted and Understandable  Dan Jensen  Director, Army Data Center Fairfield</li> </ol>	E51-345
	<p><b>Session 2B</b> IQ and Data Commercial Software  Session Chair: Linda Kresl, MIT Information Quality Program</p> <p>Commercial Off-the-Shelf Data Quality Solutions ..... pg. 228  Salomon de Jager  Chief Executive Officer  The Pilog International Group of Companies</p>	E51-361
	<p><b>Session 2C</b> Master Data Management  Session Chair: Carol Chapman, American National Insurance Co.</p> <ol style="list-style-type: none"> <li>1. Business Value of MDM and High Quality Data -- A Practical Framework ..... pg. 229  Kimmo Kontra  Manager, Information Management Services  Data Management &amp; Architecture Doman  Accenture</li> <li>2. Essential Elements of a Master Data Management Architecture ..... pg. 241  Joseph M. Bugajski  Burton Group</li> </ol>	E51-372

	<p><b>Session 2D</b> IQ in the Enterprise  Session Chair: Peter Anlyan, Anlyan Consulting, LLC</p> <p>1. Sustaining Data Quality -- Lessons from the Field ..... pg. 254  Tony O'Brien  Finance Manager  Remploy</p> <p>2. Enterprise Architecture Led Information Quality Strategy: "Two Sides of the Same Coin" ..... pg. 267  Jay Barua  Enterprise Data Architect  REI, Inc.</p>	E51-376
--	--	---------

6:15-8:45 PM	<p>SYMPOSIUM DINNER AT MIT FACULTY CLUB</p> <p><i>This dinner event is sponsored by <b>Burton Group</b></i></p>	
--------------	---	--

**THURSDAY, July 15, 2010**

8:00-9:00 AM	CONTINENTAL BREAKFAST	
--------------	-----------------------	--

9:00-10:30 AM	<p><b>Session 3A</b> NATO IQ Practices  Session Chair: Steve Arnett, US National Codification Bureau</p> <p>1. Data Quality at UK National Codification Bureau (UKNCB)  George Bond  NATO</p> <p>2. Master Data Quality (ISO 8000) Applied To Unite NATO and Other Codified Data in ERP Systems for Government ..... pg. 268  Salomon de Jager  Chief Executive Officer  The Pilog International Group of Companies</p>	E51-345
---------------	---	---------

	<p><b>Session 3B</b> IQ in Medical Informatics II  Session Chair: Bruce Davidson, Cedars-Sinai Health System</p> <p>1. caCore Evolves: Semantics II-Future Innovations  Denise Warzel  National Cancer Institute</p> <p>2. Using What You Have to Build What You Need ..... pg. 269  Patricia Gunter, MS  Senior Informaticist, Biomedical Informatics  Duke Translational Medicine Institute</p>	E51-361
--	---	---------

	<p><b>Session 3C</b> IQ Solutions Session Chair: Joseph M. Bugajski, Burton Group</p> <p>1. Business Impact of Social Media ..... pg. 287 Krish Krishnan Analyst Sixth Sense Advisors, Inc</p>	E51-372
	<p><b>Session 3D</b> Decision Modeling and Business Rules Session Chair: Charles McKinney, Freddie Mac</p> <p>1. The Decision Model: Linking Business and Technology ..... pg. 307 Larry Goldberg Managing Partner Knowledge Partners International</p> <p>2. Managing Business Rules and Information Quality across an Extended Enterprise ..... pg. 308 Mark J. Pettit Vice President, Operations Shared Services Freddie Mac</p>	E51-376
10:30-11:00 AM	COFFEE BREAK	
11 AM-12:30 PM	<p><b>Session 4A</b> IQ in the Government Session Chair: Willa Pickering, Lockheed Martin Corporation</p> <p>1. Data.gov: Answering the Demand for Open Government Data ..... pg. 317 Suzanne Acar Federal Bureau of Investigation</p> <p>2. Federal Agency Implementation of the Federal Information Quality Act ..... pg. 336 Richard B. Belzer, Ph.D. President Regulatory Checkbook</p>	E51-345
	<p><b>Session 4B</b> IQ Practices in Health Systems Session Chair: Bruce Davidson, Cedars-Sinai Health System</p> <p>1. Data Quality Assessment Framework ..... pg. 337 Laura Sebastian-Coleman, Ph.D. Ingenix, a subsidiary of UnitedHealth Group (UHG)</p> <p>2. Data Quality Challenges in Developing a Large Scale Primary Care Medical Home Pilot Project Michael Nix Fletcher Allen Health Care Randy Messier Fletcher Allen Health Care</p>	E51-361

	<p><b>Session 4C</b> IQ Management Processes  Session Chair: Joseph M. Bugajski, Burton Group</p> <p>1. Data as an Asset: Balancing the Data Ecosystem ..... pg. 348  Tonie Leatherberry  Principal  Deloitte Consulting LLP  Rena Mears  Partner  Deloitte Consulting LLP</p> <p>2. Data Quality: When is Good Enough, Enough? ..... pg. 365  Phil Teplitzky  Independent Consultant</p>	E51-372
	<p><b>Session 4D</b> Financial Industry Data Process Management  Session Chair: Charles McKinney, Freddie Mac</p> <p>1. Integrating Information Management into IT Planning ..... pg. 374  Diane Schmidt  Vice President  Freddie Mac</p> <p>2. The Federal Reserve System's Enterprise Information Management Program ..... pg. 375  Atulya Tiwari  Senior Enterprise Architect  Federal Reserve System</p>	E51-376
12:30-2:00PM	<p>LUNCH &amp; KEYNOTE SPEECH</p> <p><b>Keynote Speech</b></p> <p><i>Information Quality and the Financial Crisis</i> ..... pg. 376</p> <p><b>John Bottega</b>  Vice President and Chief Data Officer  Federal Reserve Bank of New York</p>	E51-345
2:00-2:30 PM	COFFEE BREAK	
2:30-4:00 PM	<p><b>Session 5A</b> IQ Practice in Government Agencies  Session Chair: Suzanne Acar, Federal Bureau of Investigation</p> <p>1. Information Quality at the Transportation Security Administration ..... pg. 377  Jim Watts  Program Manager  Performance Management Information System (RMIS)</p>	E51-345

	<p><b>Session 5B</b> Health Information Exchange  Session Chair: Bruce Davidson, Cedars-Sinai Health System</p> <p>1. Metadata Management in Biomedical Research: A Comprehensive Paradigm ..... pg. 391  Neera Bhansali  Department of Biomedical Informatics  Moffitt Cancer Center &amp; Research Institute</p> <p>2. My Diabetes University Application for Teens ..... pg. 402  Eugene Kolker  Chief Data Officer  Seattle Children's Hospital</p>	<p>E51-361</p>
	<p><b>Session 5C</b> IQ Models  Session Chair: Joseph M. Bugajski, Burton Group</p> <p>1. Application of Similarity-based Clustering to Entity Resolution ..... pg. 409  Thomas Schweiger  Acxiom Corporation  Xiaowei Xu  University of Arkansas at Little Rock</p> <p>2. Organized Chaos: A Framework for Classifying Data Quality Problems..... pg. 410  Joe Maguire  Burton Group</p>	<p>E51-372</p>
	<p><b>Session 5D</b> New Data Management Requirements for the Financial Industry  Session Chair: Michael Atkin, Enterprise Data Management Council</p> <p>1. Financial Industry Changing U.S. Regulatory Framework ..... pg. 431  Joseph Langsam  Managing Director, Analytic Research  Morgan Stanley</p> <p>2. Financial Industry Changing European Regulatory Framework  Francis Gross  European Central Bank</p> <p>3. Mortgage Industry Data Standards ..... pg. 432  Ted Adams  Information Management Consultant  Enterprise Data Strategy and Oversight  Freddie Mac</p>	<p>E51-376</p>
<p>4:00-4:30 PM</p>	<p>COFFEE BREAK</p>	

4:30-6:00 PM	<p><b>Session 6A</b> IQ Practice in Government Agencies  Session Chair: Suzanne Acar, Federal Bureau of Investigation</p> <p>1. Data Profiling for DQ Assessment: Rapidly Building the Business Case for DQ Improvement ..... pg. 433  David Loshin  President  Knowledge Integrity, Inc.</p> <p>2. Experiences in Data Quality ..... pg. 434  Annette Pence  The MITRE Corporation</p>	E51-345
	<p><b>Session 6B</b> Data Quality Management in Healthcare  Session Chair: Bruce Davidson, Cedars-Sinai Health System</p> <p>1. Information Quality Research from the Healthcare Perspective ..... pg. 451  Marilou Haines  President  Alamo Software, Inc.</p> <p>2. Electronic Patient Records (EPR) and Public Reporting: Benefits and Challenges to Data Quality ..... pg. 463  Elisa L. Horbatuk, MA  Data Manager, Decision Support Services  Stony Brook University Medical Center (SBUMC)</p>	E51-361
	<p><b>Session 6C</b> IQ Technology  Session Chair: Joseph M. Bugajski, Burton Group</p> <p>1. Business Insight through Real-time Analysis on Data in Motion ..... pg. 490  Marcus C. Collins  Senior Analyst  Burton Group</p> <p>2. Portfolio Approach to Control Investment on Information Quality: Case Installed Base ..... pg. 491  Jukka Koivuniemi  Modultek Oy</p>	E51-372
	<p><b>Session 6D</b> Financial Data Supply Chain Management  Session Chair: Charles McKinney, Freddie Mac</p> <p>1. Framing the Necessary Conditions ..... pg. 492  Michael Atkin  Enterprise Data Management Council</p> <p>2. Reference Data Utility Concept  Francis Gross  European Central Bank</p> <p>3. Financial Industry Reference Data Proof of Concept ..... pg. 493  Francis Parr  IBM Research</p>	E51-376

6:00-7:30 PM	SYMPOSIUM RECEPTION AND SOCIAL AT MIT FACULTY CLUB	
--------------	---	--

**FRIDAY, J**

8:00-9:00 AM	CONTINENTAL BREAKFAST	
9:00-10:30 AM	<p><b>Session 7A</b> IQ Practices in the Government Session Chair: Willa Pickering, Lockheed Martin Corporation</p> <p>1. Police Data as a "Worst Case" Example of IQ Issues ..... pg. 494 Daniel Bibel Program Manager, Crime Reporting Unit Commonwealth Fusion Center Massachusetts State Police</p> <p>2. Building Records Management into Business Processes to Improve Information Quality ..... pg. 495 Laurence Brewer Director, Life Cycle Management Division National Archives and Records Administration (NARA)</p>	E51-345
	<p><b>Session 7B</b> IQ Trends for High Tech Session Chair: Linda Kresl, MIT Information Quality Program</p> <p>1. Yahoo! Data Quality: Embedded DQ Strategy, Statistical Influences, and Challenges in Yahoo!'s Massive Data Environment ..... pg. 508 Jeff Kibler Data Quality Lead, Audience Management &amp; Analytics Yahoo! Oleksiy Chayka Yahoo! / University of Trento, Italy</p>	E51-361
	<p><b>Session 7C</b> Information Supply Chains Session Chair: Bruce Davidson, Cedars-Sinai Health System</p> <p>1. Correcting the Misconceptions About the Nature of Data that Thwart Information Quality ..... pg. 529 Lyn Robison Research Director Burton Group</p> <p>2. Accelerating Business Insight through Information Supply Chains ..... pg. 549 Sebastian Gass Enterprise Architect Chevron</p>	E51-372

	<p><b>Session 7D</b> Panel on Financial Data Management Implementation  Pathway  Session Chair: Charles McKinney, Freddie Mac</p> <p>Charles McKinney  Freddie Mac  John Bottega  Federal Reserve Bank of New York  Joseph Langsam  Morgan Stanley  Francis Gross  European Central Bank  Francis Parr  IBM Research  Ted Adams  Freddie Mac</p>	E51-376
10:30-11:00 AM	COFFEE BREAK	
11:00-12:00 PM	<p>CLOSING PLENARY SESSION</p> <p><b>Roundtable Discussion</b></p> <p>MIT IQIS in 2015: Where Do We Go from Here?  Co-Moderated by  John Talburt and Linda Kresl</p> <p><b>Closing Remarks</b></p> <p>Dr. Richard Wang  Office of the Army CIO/G6 and MIT IQ Program</p>	E51-345
	END OF THE SYMPOSIUM	