18th International Conference on Information Quality

(ICIQ 2013)

Co-located with the IAIDQ Information Data Quality (IDQ) Conference and UALR Emerging Analytics Center

Little Rock, Arkansas, USA
7 - 9 November 2013

ISBN: 978-1-62993-844-8

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571

Some format issues inherent in the e-media version may also appear in this print version.

Printed by Curran Associates, Inc. (2014)

MIT Information Quality Program
Attn: Dr. Richard Wang
225 Crafts Road
Chestnut Hill MA 02467

Phone: (617) 304-3120

rwang@mit.edu

Additional copies of this publication are available from:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: 845-758-0400
Fax: 845-758-2634
Email: curran@proceedings.com
Web: www.proceedings.com

Research Presentations

Parallel Session 1-A: Entity Resolution and IQ **Friday: November 8, 2013** **10:45 am to 12:15 pm**	**EIT 220** **Session Chair: Valerie Sessions**

A False Positive Review Indicator For Entity Resolution Systems Using Boolean Rules 1
Daniel Pullen, Pei Wang, John Talburt, and Ningning Wu

Abstract: The clerical review of potential false positive and false negative resolution decisions is critical to improving the accuracy of an entity resolution (ER) process. In ER systems using scoring rules or agreement/disagreement patterns, review indicators are triggered as pairs of references are compared. In systems using scoring rules match pairs that may need clerical review can be indicated by match scores falling within a particular value range. For systems using agreement/disagreement patterns match pairs that satisfy a particular agreement/disagreement pattern are selected for review. ER systems using Boolean match rules require a different approach. This paper describes a new method for identifying potential false positive resolutions made by an ER process based on the entropy of the identity attribute values across the complete set of references that have been linked together by the process. The method has the advantage that it can be applied to any ER process outcome regardless of the type of match rules used by the process. The method is efficient in identifying false positives in large data sets and has been implemented and tested for student enrollment data. The paper also discusses an analysis for estimating the precision and recall of false positive resolutions for various entropy value thresholds.

On Choosing Thresholds for Duplicate Detection 12
Uwe Draisbach and Felix Naumann

Abstract: Duplicate detection, i.e., the discovery of records that refer to the same real-world entity, is a task that usually depends on multiple input parameters by an expert. Most notably, an expert must specify some similarity measure and some threshold that declares duplicity for record pairs if their similarity surpasses it. Both are typically developed in a trial-and-error based manner with a given (sample) dataset.

We posit that the similarity measure largely depends on the nature of the data and its contained errors that cause the duplicates, but that the threshold largely depends on the size of the dataset it was tested on. In consequence, configurations of duplicate detection runs work well on the test dataset, but perform worse if the size of the dataset changes. This weakness is due to the transitive nature of duplicity: In larger datasets transitivity can cause more records to enter a duplicate cluster than intended. We analyze this interesting effect extensively on four popular test datasets using different duplicate detection algorithms and report on our observations.

An Approach Using Relational Markov Model For Estimating and Replacing Missing Categorical Data 21
Jianjun Cao, Xingchun Diao, Yongping Xu, and Zhen Yuan

Abstract: In order to process missing data, we propose an approach based on the relational Markov model (RMM) for estimating and replacing missing categorical data. First, for a given data set, all categorical attributes are classified as a proper number of groups, and these groups are independent of each other. Second, principles for ordering attributes in one group are proposed and the attribute sequence of the group could be indexed by the principles. Third, a RMM for estimating missing categorical value is represented. According to complete record samples, probabilities of missing value belonging to each possible value are estimated by the model. The dynamic attribute selection (DAS) method is used to select the best attribute group to estimate the missing value. The missing values could be replaced using maximum posterior probability (MaxPost) or probability proportional (ProProp) method. Finally, the effectiveness and advantage of the approach is tested by the comparative experiments on open datasets.

Research Presentations

Parallel Session 1-B: IQ in Organizations **Friday: November 8, 2013** **10:45 am to 12:15 pm**	**EIT 219** **Session Chair: C. Lwanga Yonke**

An Industry Comparison Of Data And Information Quality Priorities And Practices 39
Elizabeth Pierce and Bruce N. Davidson

Abstract: This work expands upon the analysis conducted for the joint IAIDQ-UALR IQ 2012 Industry Survey and Report on the State of Information and Data Quality [1]. In this paper we examine the differences and similarities between industries when it comes to priorities, managerial approaches, practices, tools, and maturity levels for information and data quality efforts. These comparisons raise interesting questions about how information and data quality efforts are influenced by the type of industry associated with an enterprise.

Does The Organization Size Matter? An Investigation Into IQ Effort in Accounting Information Systems Adoption 57
Andy Koronios, Wongsim Manirath, and Jing Gao

Abstract: Many organizations have undertaken information quality improvement programs and projects. In order for an organization to better target their IQ efforts, this research has conducted 10 case studies to study how IQ is managed through the accounting information system adoption process. A special focus is placed on determining how organization size influences the information quality practices. The finding is especially useful to Small and Medium Enterprises (SMEs) as many SMEs have the desire to grow bigger. By better dealing with IQ issues, there could be a successful future.

Modeling and Simulating the Impact of Social Issues on Information Quality: Literature Review 70
Therese L. Williams, David Becker, Thomas C. Redman, Amit Saha, Kashif Mehdi, Joanne Reilley, Huzaifa Syed, Wright A. Nodine, Jr., and John Talburt

Abstract: Information quality is generally defined in terms of fitness for use. Almost all agree that they prefer high-quality to low-quality information. And, while many organizations have made good progress, many find that setting up information quality programs and making improvements proves difficult. Further, most agree that the most critical difficulties stem from organizational, structural and political issues. As yet, there is no body of theory and practice to help leaders and organizations systematically understand and address these issues.

This research program aims to (begin to) build the body of needed theory. The basic idea is to employ systems dynamics and computer simulation to explore the ways hundreds of possible factors and managerial actions advance or hinder information quality efforts. More specifically then, the long-term goal of this research is to create and utilize a test bed (or simulator) to examine, in a systematic fashion, the impact of various social/cultural issues which influence the penetration and overall success of information quality in an organization. In particular, building on the work of Falleta (1), this research is a literature review of multiple organizational change models that can potentially be utilized for this modeling.

This paper reports on one aspect of this research, namely, the literature review. As one might suspect, there is much relevant work, from the fields of systems dynamics, organizational analyses, force-field analyses, and change management.

Research Presentations

Parallel Session 2-A: IQ of Wireless Sensor Networks **Friday: November 8, 2013** **1:15 pm to 2:45 pm**	**EIT 220** **Session Chair: Michael Mielke**

Big Data Quality Case Study Preliminary Findings: Hyperspectral Imaging (HSI) Using The Airborne Visible / Infrared Imaging Spectrometer (AVIRIS) 86

Dave Becker, Trish Dunn King, Bill McMullen, and Dr. Ahmed Fahsi

Abstract: In this study, we examined Big Data Quality issues using the case of the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS), hyperspectral imaging (HSI) sensor. The study addresses several factors affecting Big Data Quality at multiple levels, including collection, processing, and storage. Though not unexpected, the key findings of this study reinforce that the primary factors affecting Big Data reside in the limitations and complexities involved with handling Big Data while maintaining its integrity. For example, with the HSI imaging data, the composite, end-to-end scientific data pipeline interactions can affect its data quality. These concerns are of a higher magnitude than the provenance of the data, the processing, and the tools used to prepare, manipulate, and store the data.

Semantic-based Detection of Segment Outliers and Unusual Events for Wireless Sensor Networks 102

Lianli Gao, Michael Bruenig, and Jane Hunter

Abstract: Environmental scientists have increasingly been deploying wireless sensor networks to capture valuable data that measures and records precise information about our environment. One of the major challenges associated with wireless sensor networks is the quality of the data – and more specifically the detection of segment outliers and unusual events. Most previous research has focused on detecting outliers that are errors that are caused by unreliable sensors and sensor nodes. However, there is an urgent need for the development of new tools capable of identifying, tagging and visualizing erroneous segment outliers and unusual events from sensor data streams. In this paper, we present a SOUE-Detector (Segment Outlier and Unusual Event-Detector) system for wireless sensor networks that combines statistical analyses using Dynamic Time Warping (DTW) with domain expert knowledge (captured via an ontology and semantic inferencing rules). The resulting Web portal enables scientist to efficiently search across a collection of wireless sensor data streams and identify, retrieve and display segment outliers (both erroneous and genuine) within the data streams. In this paper, we firstly describe the detection algorithms, the implementation details and the functionality of the SOUE-Detector system. Secondly we evaluate our approach using data that comprises sensor observations collected from a sensor network deployed in the Springbrook National Park in Queensland, Australia. The experimental results show that the SOUE-Detector can efficiently detect segment outliers and unusual events with high levels of precision and recall.

Research Presentations

Parallel Session 2-B: IQ in Healthcare **Friday: November 8, 2013** **1:15 pm to 2:45 pm**	**EIT 219** **Session Chair: Bruce Davidson**

Monitoring the Quality of Clinical Administrative Health Data to Support Health System Funding 120

Maureen Kelly and Chrissy Willemse

Abstract: This presentation shares new data quality monitoring tools that were developed by the Canadian Institute for Health Information (CIHI) and the Ontario Ministry of Health and Long-Term Care to support the use of healthcare data for funding purposes.

Methods: Clinical administrative data are now being used to determine funding allocations to regions and hospitals across multiple health sectors using clinical administrative health databases maintained at CIHI. CIHI worked with the Ontario ministry to develop tools to monitor clinical data quality indicators for health care organizations on a quarterly basis, focusing on data that plays an important role in funding or that were found to be problematic through data analysis.

The tools provide a new way of looking at data quality. Data are examined at an aggregate level using graphical representation to enable quick identification of outliers. Charts automatically populate for a selected organization enabling comparisons with others and trending over time. Data visualization is used to highlight changes over time and hospital-level differences making it easy for users to identify potential data quality problems. Interpretation notes are included in the report to guide the user and encourage them to consider whether the data is an accurate representation of their patient populations and hospital processes. Questions are also posed for the user to consider whether any changes or anomalies in their data are a result of a data quality initiative or known problems.

Results and Discussion: These data quality monitoring tools foster a culture of data quality by enabling healthcare providers and health system managers to identify data quality problems and take action on an ongoing basis. This provides confidence to health system planners that the data are fit for use for funding and management purposes.

Information Quality and Data Governance for eHealth In the Era of Big Data 158

Ying Su, Cheng Dong, Marc Lange, Junpping Zhao, and Ping Yu

Abstract: This paper focuses on the second class of Big Data, and in particular on secondary uses of health care datasets in the National Health and Family Planning Commission (NHFPC) in China. For the purposes of discussion, the primary use of health care information is for the diagnosis, treatment and care that we receive from doctors, nurses and other clinicians. We ask two questions - why information quality is essential to the eHealth in the era of Big Data, and how to govern health big data? We suggest that health big data analytic engine and big data governance tool for high knowledge creation and personalized health service. We also conclude that, high quality health information about the user will lead to better personalized services, and better adaptive services.

Research Presentations

Parallel Session 3-A: IQ and Risk Management **Friday: November 8, 2013** **3:00 pm to 4:30 pm**	**EIT 220** **Session Chair: Ismael Caballero**

Total Information Quality Risk Management: Quantifying the Business Impact of Information Quality 171

Alexander Borek

Abstract: A vast body of empirical evidence has been collected which shows that using information of poor quality can cause significant risks in the organization. This practice- oriented talk presents a methodology for managing and reducing the risks inherent in information quality in form of a process for "Total Information Risk Management" (TIRM). The TIRM process has been developed in a rigorous design science cycle, which included the application of the process in six industrial in-depth case studies using participative and non-participative research and evaluation interviews with ten information management experts. The TIRM process was refined after each application using the feedback and insights collected in these studies. The studies have shown that a risk based approach can be used to quantify the business impact of information quality, which can drive the development of a robust business case for information quality initiatives.

Methods for Adjusting Expected Value of Information (EVPI) Under Situations of Data Missing Not at Random (MNAR) 189

Valerie Sessions and Stan Perrine

Abstract: Decision making under uncertainty is an extensive research field concerned with aiding the decision maker through uncertain problem spaces such as financial markets, product analysis, or medical treatment options. It is often helpful in this type of problem space to obtain additional data before making a risky or costly decision. The benefit of this new data, however, must be weighed against the cost of obtaining it. One can use the Expected Value of Perfect Information (EVPI) calculation to determine the expected payoff for receiving new information regarding a future decision. The EVPI calculation is used to place an upper bound on the amount of funding or other resources that should be spent to 'firm up' data related to an uncertain situation. Used traditionally in business analysis, this method is becoming more pervasive in medical research, education, and criminal justice. While EVPI is a useful estimate of data utility, we postulate that under cases of poor data quality, specifically data missing not at random (MNAR), EVPI calculations can be misleading without further clarification. We seek here to present the possible effects of data MNAR in the EVPI calculation. We then give examples of EVPI usage in medical literature and the reporting of sample size, missing data, and mitigation strategies used (such as list wise deletion or multiple imputation). Finally, we propose two methods that can be used to inform a decision maker of the possible effects of presumed MNAR data into the EVPI calculation. Future research shall focus on evaluating this method more fully in data sets where MNAR data is suspected.

Show Us Your Pay Stub: Income Verification In P2P Lending 201

Authors Irit Askira Gelman and Aimee A. Askira

Abstract: Peer-to-peer lending is an alternative credit market that allows individual borrowers and lenders—people like you and me—to engage in credit transactions without traditional banking intermediaries. This research centers on income verification practices in peer-to-peer lending. We report on a descriptive analysis of all the loans that were funded through Lending Club, currently the world's largest peer-to-peer lending platform, with issue dates before September 1, 2012. The score that Lending Club assigns to a requested loan is purported to encapsulate all the information that is needed for the lender to assess the risk of a potential default. This study points, however, to a potential weakness of Lending Club's loan assessment tools, which indicates that information about a loan's income verification status is in fact relevant and has value. Given this understanding, lenders' choices are surprising. Lenders that are registered directly on Lending Club's platform, including a crowd of small investors, fund a higher percentage of the listed loan amount when the borrower's income is not verified, while all other investors display the opposite, traditional risk-averse behavior.

Research Presentations

Parallel Session 3-B: Healthcare and IQ Round Table **Friday: November 8, 2013** **3:00 pm to 4:30 pm**	**EIT 219** **Session Chair: Beverly Kahn**

Healthcare and IQ Issues: A Round Table Discussion Led By

Dr. Bruce Davidson
Vice President, Performance Improvement, Hoag Health System

Dr. Alein Chun
Interim Director, Resource & Outcome Management at Cedars-Sinai Health System

Dr. Meredith Nahm
Associate Director for Clinical Research Informatics, DTMI Biomedical Informatics Core, Duke University

Mr. Sami Laine
Department of Computer Science and Engineering, Aalto University

Abstract: Information quality is an increasingly critical topic for the healthcare industry. This panel will lead an interactive discussion with participants on a variety of data quality issues associated with Electronic Health Records, Research Reporting, and Healthcare System Data. Questions to be discussed include:

1. Does the quality of healthcare information matter to clinical decisions, to facility management, and to secondary data users
2. What are all if the information uses and the IQ dimensions important to reach?
3. Will clinicians use information provided by / charted by others and under what conditions?
4. How can we make information processes transparent and how do we leverage this metadata?
5. What metadata is important in healthcare?
6. Can we improve information quality by standardizing data elements for data generate and used in care/wellness management, i.e., can the data elements be standardized, will they be accepted and will they be adopted in EHRs
7. How to assess quality, particularly accuracy of health IQ?

Research Presentations

Parallel Session 4-A: Tools for IQ **Saturday: November 9, 2013** **10:45 am to 12:15 pm**	**EIT 220** **Session Chair: Jing Gao**

Data Profiling Challenges in Engineering Asset Management Data – Conceptual Design for Next Generation Data Profiling Software 214

Jing Gao, Philip Woodall, Andy Koronios, and Ajith Kumar Parlikad

Abstract: Engineering asset management (EAM) is the process of managing the assets (from manufacturing machines to trains, planes and road bridges etc) in an organization. In order to manage these assets organizations must have good quality data about the assets. Otherwise, decisions about when to maintain an asset can be made incorrectly, and as a consequence, can adversely impact the business financially. To improve data, the first commonly accepted stage is data quality assessment, and to support this stage, data profiling software is often used. Data profiling tools can be used to uncover and measure the scale of the data quality problems and they do this by defining data quality rules. This research investigated the data profiling needs of EAM. In particular, existing profiling tools often contain generic data quality rules that are not always applicable to EAM business users. Creating EAM data quality rules without the relevant domain knowledge is very difficult and hence the best people to develop these rules are the EAM business users. This research therefore proposes an enhanced data profiling solution, which is based on the community-based central pseudo-code DQ rule repository. The proposed data profiling solution enables business users to develop and share EAM-related data quality rules promoting rule adaptability and reusability.

Solution Architectures For Retaining Data Quality Problems in Automatically Generated Test Data 225

Martin Oberhofer, Philip Woodall, and Alexander Borek

Abstract: Many organizations store sensitive data that should not be released. However, organizations often want to release this data to benefit from outsourcing of work or using the cloud for Data Quality (DQ) related tasks like data cleansing, for example. Our previous work identified useful "data generation" methods that can modify secret data to make it releasable and also retain the original DQ problems. However, there are many different types and uses of organizational data. Our aim for this work, therefore, is to present the specific architectures for data generation methods that are applicable to organizations for their different types and uses of data.

Systematic ETL Management – Experiences With High-Level Operators 236

Alexander Albrecht and Felix Naumann

Abstract: Large organizations load much of their data into data warehouses for subsequent querying, analysis, and data mining. Extract-Transform-Load (ETL) workflows populate those data warehouses with data from various data sources by specifying and executing a set of transformations forming a directed acyclic transformation graph (DAG). Over time, hundreds of individual ETL workflows evolve as new sources and new requirements are integrated continuously into the system. Managing these, often complex, ETL workflows is a daunting task.

We built an ETL management framework to improve this difficult task by providing high-level operations, such as searching, matching, or merging ETL workflows. In this paper, we present our lessons learned throughout the implementation of a prototypical ETL management framework. We discuss our observations and experiences and highlight selected suggestions and algorithms, which we propose to be suitable for building useful ETL management operators.

Research Presentations

Parallel Session 4-B: Case Studies in IQ Saturday: November 9, 2013 10:45 pm to 12:15 pm	EIT 219 Session Chair: Mariofanna Milanova

A Study of the Promotion of Information Sharing Through Presentation of Suppositional Context and Using A Concept of Corporate Household: Case Study of Tabio Corp in Japan
Hiroshi Koga ***251***

Abstract: The author proposes to use two concepts, suppositional context and corporate household, to study how the Tabio Corporation in Japan (TCJ), one of good players in the Japanese hosiery industry, improved quality of information related to POS (point of sales) data which is provided for its socks suppliers and raw material provider (The trading company which handles raw thread and dye factories) to those socks suppliers. Through a sociological analysis, the author first shows that TCJ improved the users' perspective on how to utilize the data provided (suppositional context), not by modifying the information system but by introducing the "voluntary delivery system" of socks to the distribution center. Second, the author shows that significance of entity integration (corporate household) from a view point of the socks suppliers. Finally, the author focuses on the system expansion made by TCJ to provide POS data of each shop for raw material provider. That is, the POS data was aggregated into quantities of users' products, then supplied to the users.

The Application of the IQMM Model to Evaluating the Science and Technology Information Resource Sharing Projects in China 263
Lirong Song and Xiaohong Zhang

Abstract: The problems associated with Information Quality (IQ) have become prominent restraints for China's efforts for promoting science and technology (S&T) information resource sharing. It is important to use comprehensive, authentic and accurate means to describe, measure, and evaluate the Information Quality Management (IQM) of S&T information resource sharing projects. In this paper, a framework of Information Quality Management Maturity (IQMM) assessment is proposed, which is focused on the maturity features, constituent elements, key links and processes and evaluation criteria and methods of maturity levels of IQMM to provide effective strategies and measures to improve IQ in the S&T information resource-sharing construction. Finally, the paper presents the preliminary results of a case study, where IQMM assessment is applied to some of the specific S&T information resource sharing management projects in the national program "National Science and Technology Foundation Platform" in China, to improve IQ continuously.

Improving Customer Complaint Mining 275
Matt Brown

Abstract: This presentation details the use of naïve Bayes classification to address information quality issues in a customer complaint database from a large food manufacturer. The complaint database is primarily used to identify potential problems in products, production facilities, or retail locations. The database contains information on over one million unique customer complaints and grows on average by several hundred complaints daily, as customers contact the company with complaints. The verbatim complaint is manually classified into one of numerous predetermined complaint categories. Based on this classification, the database is monitored for significant changes in the rate of complaints using Statistical Process Control (SPC) based data mining techniques. The information quality issue arises when call center operators, who often are pressed for time between calls, misclassify complaints into "unknown" or "miscellaneous categories". Subsequently, incomplete or inaccurate data limits the effectiveness of the SPC data mining technique. This paper examines the effectiveness of naïve Bayes classification in correcting these misclassified complaints.

Research Presentations

Parallel Session 5-A: IQ Dimensions **Saturday: November 9, 2013** **1:15 pm to 2:45 pm**	**EIT 220** **Session Chair: John Talburt**

Perception Of Value-Added Through A Visual Join Operation 301
Ahmed Abuhalimeh, Daniel Pullen, and M. Eduard Tudoreanu

Abstract: The quality of data and information can be judged and improved via multiple dimensions, such as degree of accuracy, degree of uncertainty, or the amount of value-added. Value-added, the focus of this paper, is one of the contextual information quality dimensions that depends on the nature of task and plays a role in the operational fitness and the goals to be achieved as ascertained by an end-user, the final information consumer. This paper presents an empirical study of how people perceive the value-added of data undergoing a join operation, which is common in both data processing and visual information fusion. The study focuses on data that is conveyed to end-users through visual representations, because graphical approaches are increasingly employed to convey large amounts of information. Two types of visualizations are shown to users, one with the basic, unprocessed data, and the other with the result of the join operation. Results show that there is an actual value added by the join operation, and we estimate it at between 24% and 35% in the context of the experiment described. This study is part of related research on the human perception of information quality conveyed through graphical means.

User Interaction Metadata for Improved Information Traceability 315
Sami Laine, Marko Nieminen, and Mika Helenius

Abstract: The origin of information must be traceable to determine its true contextual meaning and actual quality. Information flows can be traced from two complementary perspectives: information management and software systems. In management practices, various modeling methods are used to document and analyze information flows. In software systems, data lineage capability provides automatic traceability of information products back to their original data sources. However, accuracy of information is affected by many contextual and human factors that are unrecognizable from technical data flows or abstract management models. In this study, we suggest that information traceability methods should cover more metadata on such factors.

We use design science research approach to gradually develop three iterations of artifacts: theoretical categorization, laboratory prototype, and software application. In the first iteration, we gathered empirical data about user interaction situations to illustrate the potential use cases for user interaction metadata. We also categorized user interaction metadata properties that could be collected automatically from computerized processes, workflows, and user interfaces. We mapped user interaction metadata to five categories: interaction context, documentation, input controls, value properties, and interaction structures. Results from empirical cases suggest that additional user interaction metadata could help to recognize, explain, and fix contextual data quality problems in real-life information production processes.

During later iterations, our research project will develop prototypes to collect user interaction metadata and then evaluate them in real organizational settings. In the future, software-based data lineage capabilities should expand their coverage to user interaction metadata. Also, managerial modeling and analysis tools could use user interaction metadata to discover, analyze, and model actual organizational processes in more detail.

Research Presentations

Parallel Session 5-B: IQ Frameworks **Saturday: November 9, 2013** **1:15 pm to 2:45 pm**	**EIT 219** **Session Chair: Elizabeth Pierce**

I8K: An Implementation of ISO 8000-1X0 331

Ismael Caballero, Isabel Bermejo, Luisa Parody, Mª Teresa Gómez López, Rafael M. Gasca, and Mario Piattini

Abstract: The exchange of master data between organizations can be regulated by the family of standards ISO 8000-1x0. This family of standards imposes several requirements for some of the activities derived from the used of master data management. These requirements are the existence of a specific syntax to be stored in a data dictionary (ISO 8000-110), the necessity to include information about the data provenance (ISO 8000-120), the necessity to include information about the assessment and certification of data quality levels for the dimensions of accuracy (ISO 8000-130) and completeness (ISO 8000-140).

The data exchanged between applications requesting data and their corresponding data providers tend to be developed by means of Web Services, and this way of communication is covered by the standard ISO 8000-1x0. In order to satisfy the requirements of the family of standards, two elements are included in this paper as the main contribution: The I8K architecture that that implements some of the requirements of the specified part of ISO 8000 with a web service approach., and the ICS-API, which provides developers with the necessary primitives to communicate the applications to use the standard with the I8K architecture. With the aim of illustrating the use of both components, we also introduce an example in the domain of travels, in which an application named TripPlanner exchanges data with a flight provider named FlightCIA.

The Information Value Methodology: How Average Users Assess IQ On The Web – Preliminary Results 346

Marilou Haines and Elizabeth Pierce

Abstract: The Internet is a self-regulating complex system where users decide what is important by their actions. Since the burden of locating and evaluating information depends on knowledge, experience and skill, this study investigates the web-user experience in a rigorous and holistic manner. The survey instrument, populated with vetted Quality Characteristics from key multi-disciplinary literature, best practices and international standards presents the point of view of academics and practitioners. Over 200 students and faculty of a large U.S. university assessed the criticality of those dimensions. The analysis of their responses will advance our knowledge on the missing factor: "the perspective of the information consumer" and serve as a baseline for future research. This paper offers preliminary results of the data collected through April 2013.

Research On Information Quality Viewed by Praxiology 361

Zbigniew J. Gackowski

Abstract: This paper discusses the current Framework for Information Quality Research from the praxiological perspective, the theory of human conduct; authors of the Framework encourage interdisciplinary approaches. The praxiological perspective reveals the need for extension of the topics of the framework with regard to how differently information quality is perceived along the line of command, when strategies and theories of operations management shift, and when one faces the dangers of disinformation, misinformation, or even outright information warfare as studied in war academies.

Research Presentations

Parallel Session 6-A: IQ Research Workshop **Saturday: November 9, 2013** **3:00 pm to 4:30 pm**	**EIT 219** **Session Chair:**

Information Quality and Supercomputing 375
Richard Segall and Qingyu Zhang

Abstract: This research pertains to the study of information quality and how this relates to the increased spectrum of dimensionalities for computation in speed, magnitude and intensity of data as obtained upon the applications of supercomputing to visual analytics and data mining.

Supercomputers are used today for highly-intensive calculation tasks for projects ranging from quantum physics, weather forecasting, molecular modeling, and physical simulations. Actual applications of supercomputers are shown in this research for 3D visualization and microarray data analysis for breast and lung cancer. Many of the open source visualization software for use by supercomputers are also discussed for the types of information quality that each of these can achieve.

IQ Challenges in Big Data N/A
Elizabeth Pierce et al

Abstract: This discussion session will focus on the IQ challenges of big data including how to incorporate these topics into undergraduate and graduate curriculum.

Emerging Markets and Collaborations for IQ N/A
John Talburt et al

Abstract: This round table will provide a forum for researchers to network with colleagues from around the world on IQ projects and investigations.