The summer is finally here, but then again it has been here all along. The US had only one week of freezing weather, instead of the normal eight to twelve, and had one of the mildest winters in recorded history. Of course, if next winter starts to look like the last one, then there could be a general panic, as we realize we truly are experiencing a change in the climate. With the last day of the Mayan calendar on Friday, December 21, 2012, it will be, by all accounts, an interesting year.

Given the growing number of books on the subject of data quality, I felt it was time to contribute my own perspective. My first book titled, “Managing Blind: A data quality and data governance vade mecum,” documents my personal journey looking for practical solutions to data quality and data governance. There were a lot of unexpected twists and turns over the years, and by documenting my own adventures I hope you will be able to do better. The book is available from ECCMA in .PDF format and will be available as an eBook or iBook from Amazon, Barnes and Noble, as well as, the Apple Store in July. Hardcopies are expected to be available in 2013.

In a blog on the latest SAP user conference SAPPHIRE 2012, it asked, “What are the common components in all of the hottest topics at SAPPHIRE”? The answer, “data quality.” Commenting on the latest business intelligence tools, the blog went on to say, “In the end, these innovative new software tools like SAP HANA can’t do what they are meant to do for your business if you have poor

(Continued on page 2)
data quality - unless your goal is simply to make your incorrect and error-ridden information look a lot prettier.” I could not agree more. If you Google, “data quality” you will get over 2 million hits and as evidenced by the advertising associated with the terms, “data quality,” at least Google is making money from it.

It is also clear that the real solution to data quality is the ability to obtain quality data from an authoritative source. This requires the ability to define requirements for data and to communicate these requirements in a way that they are easy to understand. This is why ECCMA is committed to developing open international standards, as well as, the eOTD®, eDRR™ and eGOR™ registries. The awareness of the benefits, if not the challenges of data quality, are now evident. ECCMA is making some changes to better support those who are faced with making data quality and data governance, a practical reality.

Starting July 1st, 2012, the ECCMA office in Bethlehem will be leading an international membership campaign. A key component in this campaign will be the ECCMA Corporate Dictionary Manager (eCDM™), an open source cloud application. The eCDM is designed to make it easier for ECCMA members to create and maintain their corporate dictionaries as subsets of the eOTD.

Sheron Koshy, President of ECCMA India, will be leading the development of the eCDM and his team will also be providing support for ECCMA members interested in using the new application. The eCDM will make it much easier for ECCMA members to contribute to the eOTD. There are also two important new features in the eCDM that are designed to help ECCMA members with both internal and external communication. The first is the spreadsheet, report, form and data model registries. These registries are designed to ensure that the corporate dictionary is rooted in the practical reality of everyday business. The ability to link the dictionary concepts to the headings, lists and codes used in spreadsheets, reports, forms and data models is a unique way to monitor

For more information regarding the ECCMA Corporate Dictionary Manager please visit: www.eccma.org/eCDM.
use and for the dictionary steward to encourage harmonization of terminology throughout the business. The second is concept equivalence; this is designed to make it easier to communicate with other users of the eOTD, as well as, to make it easier to take advantage of concept localizations.

A mirrored copy of the eOTD has been created using an Oracle database to allow the eCDM to be developed using Oracle Application Express (APEX), a rapid development tool for Web applications. All members of ECCMA are welcome to submit feature and functionality requests. Members who would like to learn or have experience with APEX and would like to participate in the development of the eCDM, can contact Sheron (sheron.koshy@eccma.org). He is encouraging any input.

Another important initiative for 2012, is the ECCMA Global Organization Registry (eGOR). This is a cooperative vendor master designed to support Cataloging at Source (C@S). The model for eGOR is detailed in a White Paper that is available in the “Downloads” section of our webpage (www.eccma.org/resources/downloads.php). I would encourage everyone to take a look at it because it contains some interesting data structures that can be useful. Of course the most immediate benefit, is to register your vendors in eGOR and obtain an ECCMA issued public domain organization identifier for each vendor. The ECCMA identifier is specifically designed so you can exchange vendor performance and related information with other buyers without the risk of infringing proprietary organization identifiers from commercial data providers. Also, if you are a Full Member, ECCMA will assist you in validating vendor data, as well as, obtaining missing vendor data. This is an important benefit of membership, so please take advantage of it.

Finally, the demand for ISO 8000-110:2009 Master Data Quality Manager training and certification continues to grow and a new on-line multiple choice test has been introduced for certificate renewal. If you are interested in a training course or in certification please contact Melissa (melissa.hildebrand@eccma.org).

We are looking forward to a great summer at ECCMA and hope you will join us in our efforts within the coming months!

Respectfully submitted,

Peter R. Benson, ECCMA Executive Director

For more information on becoming a Member, please visit: www.eccma.org/membership/membership.php
Update on Scope’s prOdigi™ Product Lab
Launched Early 2012

Scope eKnowledge Center (Scope), a Quatrro group company and a leading knowledge service provider, launched its product lab, prOdigi™, on February 2, 2012, to ideate, develop and deliver superior technology-based service delivery platforms aligned with customer demand, thereby building scale and maturity in existing products. The launch is closely aligned with Scope’s strategy to productize its proven service delivery capabilities across a number of service areas in content enhancement, data, intellectual property and business research to take the offerings to a wider and more diversified base of clients.

From the time of launch, prOdigi has been significantly accelerating the development and enhancement of Scope’s products such as Authentik™ and ConSCIse™ and has been responsible for the successful completion of challenging pilot projects for several industry-leading players. These products have allowed Scope to demonstrate capability for improved standardization and quality of deliverables in large-scale projects without compromise in turnaround times.

The prOdigi lab has successfully completed the enhancement of the validation module of Scope’s AuthEntik author data management solution. A dedicated team of researchers worked towards upgrading AuthEntik with an advanced standardization module that is capable of automatically mapping author data record elements with information in the thesaurus, authority file, etc. Further, the team has developed and tested the first version of an optimal disambiguation module to conduct disambiguation of author names using affiliation details and email identification.

Teams at prOdigi are presently working on the further enhancement of ConSCIse (launched late 2011) across multiple modules. prOdigi envisages releases of new versions of ConSCIse for text-rich document summarization, book summarization and press release summarization to meet the needs of diversified publishers including academic, business-to-business, business-to-consumer, trade, news and event publishing, among others. New screen designs, additional functions and advanced features for optimal digital delivery of content abstracts across media and publishing formats are coming up shortly.

Among new products due for launch is Scope’s spend data analysis solution, which is specially equipped with capabilities to solve the data management issues for small and medium business across verticals. This solution, planned to be launched in North American and European markets, has been modeled on DataMan™, Scope’s enterprise data management solution comprising de-duplication, standardization and enrichment modules, for making enterprise data buried in data warehouses/business intelligence platforms clean and ready for further analysis leading to predictive modeling and more advanced data manipulation.

prOdigi is also incubating a solution for patent classification and clustering, which is expected to bring significant speed in the implementation of IP strategies of large corporate customers, as well as, tremendous competitive advantage to Scope.

About Scope eKnowledge Center
Scope eKnowledge Center, a Quatrro Group company, is an award winning provider of outsourced, cost-effective content enhancement, data management and intellectual property and research services to the global information industry. In operation for over 24 years, Scope uses a combination of deep subject matter expertise in a number of businesses and STM domains, innovative technology and proprietary intellectual property to provide world class services. Scope is also the publisher of the Knowledgespeak daily newsletter, a comprehensive news service focusing on the STM information industry.
Tom Carlock, Dun & Bradstreet
“How to Grow your Business with Intelligent Insight”
The presentation will cover how to enhance the accuracy, currency and completeness of your data, how to develop more intelligent business insight to grow revenue and how to streamline and automate your data management processes to reduce costs.

Andrea Cifor, Microsoft
“Information Architecture: Evolving Legacy Systems from Chaos to Order”
The presentation will be a walk through on evolution of an ecosystem where the organization is faced with disconnected data sources and a need for a managed intelligent solution based on three objectives.

Alexander Edelstein, Servio
“Humans at Scale: Addressing Labor Bottlenecks in Data Quality Projects”
The presentation will be on case studies showing how Fortune 100 companies have tapped into enterprise crowd sourcing to clean and categorize data and metadata.

Ina Felsheim, SAP
“Step-by-Step Guidelines for Building an Information Governance”
The presentation will cover common pitfalls of information governance programs and how to ensure future success.

Zbigniew Gackowski, California State University Stanislaus
“Alignment of Information Quality and Risk Management”
The presentation will focus on challenges aligning information quality management to different theories of operations and shifting political, military and business strategies; how shifts in this respect may change objectives, focus, scope and priorities in quality and risk assessment.
2012 DQS² Speaker Updates

Steven Gokorsch, Rockwell Automation
“A Standards Based Approach to Multi-Language Product Descriptions”
The presentation will provide practical information on how to succeed with creating multilingual product descriptions and provide high quality, technically correct descriptive information to support global businesses.

Jim Hart, SaS
“Data Governance: The Cornerstone of MDM”
The presentation will examine how a single platform for the key requirements of data management, including data quality and data integration capabilities, can help companies fix their data problems and realize immediate benefits from their data assets.

Craig Laufer, Erie Insurance
“The Road to Noncompliance is Paved with Good Data”
The presentation will show organizations how they can meet the need for realistic test data while minimizing privacy risks, as well as, the pros and cons of various solutions.

Robin Rappaport, IRS
“Data Quality: The Journey at IRS Research, Analysis, and Statistics (RAS) from 2005 to present”
The presentation will be on the Compliance Data Warehouse (CDW) and how it aims to address data quality in 5 areas: timeliness, relevance, accuracy, interoperability and coherence.

CHECK LINKEDIN FOR ALL THE EXCLUSIVE UPDATES FOR THE 2012 DQS² ECCMA OPEN FORUM GROUP
Linking the knowledge of today with the power of tomorrow

2012 DQS² Speaker Updates

Doug Stacey, Allstate Insurance Company
“Code Management Systems in Large Enterprises”
The presentation will describe an approach for the centralized management of codes and values based on a Domain Meta-Model implemented in a custom metadata repository.

C. Lwanga Yonke, IAI/Q
“The Skills of the Information Quality Professional”
The presentation will be on a study of a broad group of information/quality practitioners across the globe to develop a consensus answer to these important questions.

Become a Sponsor at DQS²

ECCMA is offering various sponsorship opportunities at this year’s Annual Data Quality Solutions Summit, October 23rd to 25th, 2012, in Santa Barbara, California.

Sponsorship Benefits:
- Pre-marketing brochures
- E-mails
- Social media sites
- Program booklets
- Event signage
- General session slide presentations
- ECCMA newsletter
- ECCMA website

For more information on benefits please contact Melissa Hildebrand: melissa.hildebrand@eccma.org

Sponsored Days:

<table>
<thead>
<tr>
<th>Days</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Full Day</td>
<td>$3,000</td>
</tr>
<tr>
<td>2 Full Days</td>
<td>$5,000</td>
</tr>
</tbody>
</table>

Price includes recognition throughout the entire day.

Sponsored Meals:

<table>
<thead>
<tr>
<th>Meal</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breakfast (2 days)</td>
<td>$2,000</td>
</tr>
<tr>
<td>Lunch (2 days)</td>
<td>$3,400</td>
</tr>
<tr>
<td>Welcome Reception (2 hours)</td>
<td>$1,500</td>
</tr>
<tr>
<td>Fellows Award Dinner (2 1/2 hours)</td>
<td>$2,000</td>
</tr>
</tbody>
</table>

Price includes recognition for the duration of the meal.

ECCMA strives to be fair and adjust to the needs of companies interested in sponsoring and take into consideration the benefits provided.
Provenance

By: Gerald Radack, Chief Technical Officer, ECCMA

What is “provenance”? It is most certainly not a region in southeastern France. The dictionary defines “provenance” as, “the place of origin or earliest known history, especially of a work of art, manuscript, etc.” The reader of this newsletter, however, is most likely interested in data provenance. ISO 8000-2 defines the concept of “data provenance record” as a, “record of the ultimate derivation and passage of a piece of data through its various owners or custodians”\(^1\). (Note that ISO 8000 does not provide a formal definition of “data provenance,” relying instead on the generally understood meaning of “provenance,” combined with the term “data.”).

Some of the uses of provenance information include:

- assessing the reliability and timeliness of a piece of data, based on who created it originally, when it was created and who handled it since it was created;
- ensuring that legal and regulatory requirements for data handling were met;
- avoiding data echo. Data echo occurs when an organization obtains your data, possibly mixing it with other data, then sells it back. If you add such data to your master data without checking the provenance, you could be overwriting newer data with older data.

ISO 8000-110 provides a set of basic requirements for achieving data quality at the property value level. ISO/TS 8000-120 (Part 120) is an optional add-on to ISO 8000-110 that provides requirements for recording and exchanging data provenance information. In order to be able to claim that a data set conforms to Part 120, each property value must be associated with a provenance record.

(Continued on page 9)
A provenance record contains one or more provenance events. Examples of provenance events include:

- create: the property value was created;
- extract: the property value was extracted from a database;
- load: the property value was loaded to a database.

While the benefits of recording data provenance information is widely recognized, mainstream platform technologies such as databases and programming languages do not provide built-in support for capturing and manipulating provenance information. In order to avoid creating unnecessary barriers to implementation, the current (first) edition of Part 120 takes a very lightweight approach to provenance information. Just two database tables, “provenance record” and “provenance event,” are needed to record Part 120 compliant provenance information.

Part 120 does not specify a particular set of provenance events. The eOTD® Implementers Forum is currently working on a list of provenance events for master data. Following prototyping by ECCMA members, this might be proposed as an addition to a future version of the ISO 8000 standard. If your organization could benefit from the provenance information, please send me a short description of your use case(s) (gerald.radack@eccma.org). Any ECCMA member is also welcome to join the eOTD Implementers Forum. Please contact Vicky Falcone (vicky.falcone@eccma.org) to be added to the mailing list.
EDF

Every day tens of millions of businesses and consumers turn to EDF—one of the world’s leading electric utilities—for the electricity they need to keep their homes, offices and factories running smoothly.

That’s why we at EDF are committed to:

- giving you a reliable, energy-efficient power supply backed by unbeatable customer service;
- ensuring maximum energy safety;
- remaining at the forefront of energy technology—whether for nuclear, hydro, wind or solar power—and leading the drive towards a carbon-free world.

To learn more visit: www.edf.com.

Integra Consultoria

Integra has an experienced team with extensive technical knowledge of materials constantly updated to offer the customer solutions and services that exceed your expectations.

Integra is focused on generating effective results for clients, and developing new customers. Integra performs activities in a practical and objective manner to achieve the goals in the shortest time possible, while generating credibility with customers.

Integra has clients of different sizes, markets and growth stages. Therefore, it provides solutions adapted to the reality of each client, always striving for quality, compliance with the agreed deadlines and generating the expected results.

To learn more visit: www.integraestoque.com.br.
Linking the knowledge of today with the power of tomorrow

Who’s Who?  Renewed Members

**ASSOCIATE MEMBER**

**Scope eKnowledge**

Scope eKnowledge Center is a reliable partner in providing spend analysis, procurement, and supply chain services to business communities across the world. We assist our clients to profile, cleanse, standardize, enrich and classify product descriptions from both unstructured (Excel spreadsheet, other maverick spend) and structured data (i.e. product, vendor and others) and also provide follow-on value added data analysis and intelligence. Our clients benefit from access to over 500 subject experts, certified MDM, procurement and research professionals and faster turnaround times.

To learn more visit: [www.scopeknowledge.com](http://www.scopeknowledge.com).

**ASSOCIATE MEMBER**

**Sumaria Systems, Inc.**

Since its inception in 1982, Sumaria has provided a wide variety of Information Technology (IT), Engineering and Professional services to the U.S. Department of Defense (DoD) and additional government agencies. We have a proven track record of delivering cost-effective services and solutions to our clients on time and within budget. With more than 28 years of experience and hundreds of successful projects, Sumaria provides a diverse range of technology-based solutions, program management support and acquisition support services to the DoD and other government agencies. We staff technology and management professionals and attract the best and brightest personnel to support our clients. We operate with the highest degree of integrity, objectivity and competency to provide outstanding performance that creates sustained, long-term relationships with our customers. We have specialized professionals for every need and flexible contract vehicles for comprehensive services and rapid reaction solutions and support. Our streamlined structure and effective contract management ensure proactive client satisfaction.

Sumaria’s operating locations in Alabama, Colorado, Georgia, Illinois, Massachusetts, Nebraska, Ohio, Oklahoma, Utah and Virginia allow us to provide a full range of services and local capabilities to our customers, and to respond rapidly and accurately to meet our clients’ surge requirements.

To learn more visit: [www.sumariasystems.com](http://www.sumariasystems.com).
Tender.Pro is an e-procurement system on b2b market. Company Tender.Pro has been working since 2002 and has extensive experience allowing developers to create a service useful for suppliers and customers. From easy search through a huge catalog of products, through price lists through and procurement notices to the service of various types of biddings - it is all available to clients systems along with the professional support team at Tender.Pro.

To learn more visit: www.tender.pro.
About ECCMA

Formed in April 1999; the Electronic Commerce Code Management Association has brought together thousands of experts from around the world and provides them a means of working together in the fair, open and extremely fast environment of the Internet to build and maintain the global, open standard dictionaries that are used to unambiguously label information. The existence of these dictionaries of labels allows information to be passed from one computer system to another without losing meaning.

ECCMA 2980 Linden St., Ste. E2 Bethlehem, PA 18017 Tel: +1 610 861-5990 Fax: +1 610 625-4657 www.eccma.org

ISO 8000-110:2009 Master Data Quality

Upcoming Newsletter Issue

If you are a member of ECCMA in good standing, we invite you to submit articles for our viewers to read. It can be anything in the data quality industry, cataloging projects or interesting news and tips you’d like to share with our fellow members and audience. If you are interested please send an email to editor@eccma.org. Thank You!

Next Article Due By
August 24, 2012

Next Release Date
September 3, 2012