



# 13<sup>th</sup> Conference on Software Engineering and Databases

## XIII Jornadas de Ingeniería del Software y Bases de Datos

Gijón (Spain), October 7-10 2008

**EDITORS:** Ana Moreira  
María José Suárez-Cabal  
Claudio de la Riva  
Javier Tuya

# **13<sup>th</sup> Conference on Software Engineering and Databases**

# **XIII Jornadas de Ingeniería del Software y Bases de Datos**

**Gijón (Spain), October 7-10 2008**

**EDITORS:** Ana Moreira  
María José Suárez-Cabal  
Claudio de la Riva  
Javier Tuya

Edita:  
Ana Moreira  
María José Suárez-Cabal  
Claudio de la Riva  
Javier Tuya

Filmación e impresión:  
Gráficas Rigel

Depósito Legal:  
AS - 5.236 - 08

ISBN:  
978-84-612-5820-8

## **Volume Editors Details**

### **Ana Moreira**

Departamento de Informática  
Faculdade de Ciências e tecnologia  
Universidade Nova de Lisboa  
2829-516 Caparica, Portugal  
E-mail: amm@di.fct.unl.pt  
URL: <http://ctp.di.fct.unl.pt/~amm/>

### **María José Suárez-Cabal**

Departamento de Informática  
Universidad de Oviedo  
33204 Gijón, Spain  
E-mail: cabal@uniovi.es

### **Claudio de la Riva**

Departamento de Informática  
Universidad de Oviedo  
33204 Gijón, Spain  
E-mail: claudio@uniovi.es  
URL: <http://www.di.uniovi.es/~claudio/>

### **Javier Tuya**

Departamento de Informática  
Universidad de Oviedo  
33204 Gijón, Spain  
E-mail: tuya@uniovi.es  
URL: <http://www.di.uniovi.es/~tuya/>

# Preface

## Celebrating 13 Years of JISBD

With the 2008 edition in Gijón (October 7 to 10), the Conference on Software Engineering and Databases (JISBD) celebrates 13 years of existence. Born as a forum where the Spanish community would publish their work, meet to discuss potential research collaborations and evaluate the progress of research projects funded by the Spanish Ministry of Science and Technology, JISBD has long since moved beyond its initial boundaries and crossed several oceans.

Presently, the conference has become an important reference for younger researchers, as well as a forum which the more experienced do not wish to miss. In recent years, JISBD has broadened its radius, accepting papers also in English and Portuguese, in addition to Spanish. This change, not only brought more conference participants, but also significantly increased the number of submissions and, principally, the quality of the submissions accepted.

The JISBD community is now self-sustained and continues to expand. The quality of work accepted is equivalent to that of other relevant international events. In recent years, it has been possible to edit a special volume of IEEE LA with extended versions of the best conference papers and this also is happening with the current edition. This special issue, together with the conference proceedings with ISBN, is a showcase of the quality of the work of JISBD.

One of the highlights of this conference has been the excellence of its keynote speakers. Many of the most admired international researchers and professionals have already been invited to address the JISBD participants.

Within this rich framework for scientific and technological interchange, the conference includes several satellite events. In addition to the presentation of high quality original papers in the main conference, the program includes tutorials, tool demonstrations and workshops for the discussion of innovative ideas and work in progress, as well as a forum to bring to a wider audience research work already published in prestigious journals or conference proceedings (with an acceptance rate below 25% and an impact factor above 0.5).

It is no exaggeration to claim that JISBD has been consolidating its position as a reference event where researchers and professionals of Software Engineering and Databases can get together to discuss results and share ideas. JISBD has become an important forum for collaboration between different strands and research groups, while continuing to offer its participants a well organized event with exceptional hospitality.

## About this Edition

The increased global reach of JISBD is evident in the origin of papers received. This year, in addition to the two Iberian and ten Latin-American countries, submissions arrived also from China, France, Germany, India, Iran and Pakistan.

Of a total of 115 abstracts, 112 papers were submitted for review. Most papers were reviewed by three PC members, and several were reviewed by four. The program Committee accepted 30 full papers and selected 12 for presentation as short papers. The acceptance rate for full papers was approximately 25%.

The increasing success of the conference implies greater responsibilities in terms of guaranteeing independent judgement and ensuring compliance with international standards of ethics. For this reason, a greater effort has been made in recent years to avoid double submissions, a task made

more difficult by the fact that the conference accepts submissions in three languages. This year, three good papers were rejected due to double submission, in different languages to different events.

In addition to the accepted papers, the conference includes five workshops, one tutorial, nine tool demos, an industrial panel and also a forum to discuss important relevant work already published elsewhere.

A highlight of the conference is, without doubt, the excellence of the invited keynote speakers. This year is no exception and we are honoured indeed to receive Bashar Nuseibeh and Bran Selic.

Bashar Nuseibeh is an academic and researcher at the Open University in the UK and invited professor in various other universities, including Japan's National Institute of Informatics. Bashar chairs several international committees and is recognized also for industry work, including organizations such as the UK's National Air Traffic Services (NATS), Texas Instruments, Praxis Critical Systems, Philips Research Labs, and NASA.

Bran Selic was, for many years, a distinguished engineer and researcher at IBM, and currently heads a global consultancy based in Canada. He is internationally known for his work in large-scale industrial systems, and for his pioneering work in Model-Driven Development and Real-Time Embedded Systems.

Bashar's keynote is entitled "*The five W's (and one "H") of Security: ... Software Engineering of Secure Systems*" while Bran's is on "*Model-Based Software Engineering: Expected and Unexpected Challenges*".

## Acknowledgements

A very special word of thanks is due to Bashar and Bran for having accepted my invitation and for sharing all the participants their knowledge, experience and refined wit. I sincerely hope JISBD was also for them a gratifying and unique experience.

Acknowledgements are due to a multitude of collaborators without whom the conference could not have been a success. Firstly, the paper authors for the trust placed in the quality of JISBD as a conference that merited their submissions. Secondly, to the PC members, whose diligent review work ensured that the authors' trust continues to be justified.

For managing the submission and review process, I was fortunate to have the constant help of Juan Hernández and José Javier Berrocal; they were my guardian angels, constantly alert to deadlines and ready to help as necessary. A special thanks for the contribution of my "Executive Program Committee", Antonio Vallecillo, Juan Hernández, Miguel Toro, Vicente Pelechano and Xavier Franch.

Acknowledgement is due to the main conference organizers, especially to Javier Tuya and his co-chair, Claudio de la Riva, for their efficient handling of the numerous tasks that a conference of this size and quality entails. Thanks also to those responsible for the satellite events (in alphabetical order), António Rito Silva, Antonio Vallecillo, Gustavo Rossi, João Araújo, João Falcão e Cunha, José Berrocal, José Corrales, José García-Fanjul, João Miguel Fernandes, Lidia Fuentes and María José Suárez-Cabal.

Finally, a special word of thanks to the sponsors of this conference, without whose contribution the event would have been somewhat less charming (not to mention gastronomically less satisfying).

Ana Moreira  
Program Committee Chair

# Prefácio

## Celebrando 13 Anos de JISBD

Com a edição de 2008 em Gijón (7-10 Outubro), a Conferência em Engenharia de Software e Bases de Dados (JISBD) celebra 13 anos de existência. Apesar de ter nascido como um fórum onde a comunidade espanhola publicava os seus trabalhos e se reunia para discutir potenciais colaborações futuras de investigação, e até avaliar o estado de andamento dos projectos de investigação financiados pelo Ministério da Ciência e Tecnologia espanhola, há muito que extravasou essas fronteiras e cruzou oceanos.

Actualmente o JISBD é um marco importante na investigação dos mais jovens, mas também um fórum que os mais seniores não querem perder. Nos últimos anos a conferência abriu-se para o mundo inteiro, aceitando artigos escritos em Inglês, Espanhol e Português. Esta viragem trouxe não só mais participantes à conferência, mas também um aumento significativo do número de trabalhos submetidos e, principalmente, um aumento na qualidade desses trabalhos.

A comunidade do JISBD é agora auto-sustentada e em contínua expansão. A qualidade dos trabalhos aceites é equiparada à de muitos outros eventos internacionais de relevo. Por este motivo, nos últimos anos, foi-nos possível editar um volume especial no IEEE LA com uma versão estendida dos melhores trabalhos da conferência, o que acontecerá também nesta edição. Este volume, em conjunto com as actas formais da conferência com ISBN, é uma mostra da qualidade do trabalho que aqui se discute.

Uma das características de excelência desta conferência tem sido, desde sempre, o gabarito dos seus palestrantes convidados. É um prazer ver que muitos dos mais admirados investigadores e profissionais internacionais já foram convidados a falar para os participantes do JISBD.

Neste enquadramento fecundo para divulgação científica e tecnológica, a conferência inclui vários eventos satélite. Além dos artigos seleccionados para apresentação na conferência, o programa inclui ainda tutoriais, demonstrações de ferramentas, workshops para discussão de ideias inovadores e trabalhos em andamento, assim como um evento para a disseminação de trabalho de investigação já publicado em revistas e actas de conferências de grande prestígio (onde o índice de aceitação é inferior a 25% e o factor de impacto superior a 0.5).

Assim, não é excessivo afirmar que o JISBD se tem vindo a consolidar como um evento de referência onde investigadores e profissionais em Engenharia de Software e Bases de Dados se encontram para discutir, disseminar e trocar ideias, partilhar experiências e resultados entre diversos sectores e grupos de investigação, num contexto de excelente organização e invulgar hospitalidade.

## Sobre esta Edição

A atestar o crescimento e internacionalização do JISBD está a origem dos artigos que nos chegaram. Este ano, a nacionalidade dos autores foi surpreendentemente diversificada, pois para além dois países Ibéricos e de dez países Latino-Americanos, recebemos trabalhos também da Alemanha, China, França, Índia, Irão e Paquistão.

O número total de resumos foi de 115, sendo que destes, 112 artigos foram submetidos para avaliação. Cada artigo foi avaliado por pelo menos três revisores, sendo que vários foram avaliados por quatro. O Comité de Programa aceitou 30 artigos longos e escolheu 12 para apresentação como artigos curtos. Assim, o índice de aceitação de artigos longos foi de cerca de 25%.

Este sucesso acarreta responsabilidades acrescidas em garantir a independência de julgamentos e em fazer cumprir a ética e as normas internacionais. É por este motivo que, nos últimos anos, se

tem feito um esforço muito grande para evitar submissões duplicadas, tarefa nem sempre fácil para os membros do Comitê de Programa, já que a conferência aceita três línguas de escrita. Este ano foram rejeitados três bons artigos avaliados como de submissão duplicada, em duas línguas, para eventos diferentes.

Para além dos artigos seleccionados, a conferência conta também com a organização de cinco *workshops*, um *tutorial*, nove demonstrações de ferramentas, um painel industrial e ainda um fórum onde se discutem trabalhos de relevo já publicados em revistas ou outras conferências.

Mas sem dúvida que os momentos mais altos da conferência são sempre marcados pelo admirável conjunto de palestrantes convidados. Este ano tivemos a sorte de receber Bashar Nuseibeh e de Bran Selic.

Bashar Nuseibeh é um académico e investigador da Open University, na Inglaterra, e professor convidado em várias outras universidades, incluindo o Instituto Japonês de Informática. Bashar preside vários comités internacionais e é admirado também pelo seu trabalho para a indústria, que inclui organizações como o National Air Traffic Services (NATS) do Reino Unido, Texas Instruments, Praxis Critical Systems, Philips Research Labs, e a NASA.

Bran Selic foi durante umas dezenas de anos engenheiro e investigador distinguido da IBM e actualmente preside uma empresa de consultoria internacional sediada no Canadá. É conhecido mundialmente pelos seus trabalhos em sistemas de larga escala industrial e também pelo seu pioneirismo nas áreas de desenvolvimento orientado a modelos e sistemas embutidos de tempo real.

A palestra do Bashar é intitulada “*The five W’s (and one “H”) of Security: ... Software Engineering of Secure Systems*”, enquanto que a do Bran é sobre “*Model-Based Software Engineering: Expected and Unexpected Challenges*”.

## Agradecimentos

Uma palavra especial de agradecimento ao Bashar e ao Bran por terem aceite o meu convite e por brindarem todos os participantes com a sua experiência, conhecimento e refinado sentido de humor. Espero que o JISBD tenha sido também para eles uma experiência agradável e diferente.

Agradecimentos são justamente devidos ao grande número de colaboradores, sem o contributo dos quais, a conferência não poderia ter tido êxito. Aos autores, claro, por confiarem na qualidade do JISBD e submeterem, por isso, os seus trabalhos. Aos membros do Comitê de Programa cujas revisões asseguram que essa confiança continua a justificar-se.

Para gerir o sistema de submissão, contei com o apoio incondicional do Juan Hernández e do José Javier Berrocal. Eles foram os meus “anjos da guarda”, sempre atentos a todos os prazos e prontos a dar todas as explicações. Um agradecimento particular ao contributo meu “Comitê Executivo de Programa”, Antonio Vallecillo, Juan Hernández, Miguel Toro, Vicente Pelechano e Xavier Franch. Obrigada pelo vosso apoio e sugestões.

Obrigada aos organizadores principais da conferência, em especial ao Javier Tuya, e ao seu vice-presidente, Claudio de la Riva, pela gestão eficaz das inúmeras tarefas que uma conferência desta dimensão exige. Um agradecimento é ainda devido, e por ordem alfabética, aos responsáveis dos eventos satélite, António Rito Silva, Antonio Vallecillo, Gustavo Rossi, João Araújo, João Falcão e Cunha, José Berrocal, José Corrales, José García-Fanjul, João Miguel Fernandes, Lidia Fuentes e María José Suárez-Cabal.

Finalmente, um agradecimento aos patrocinadores da conferência, sem o contributo de quem o evento teria tido menos charme (e uma gastronomia muito menos requintada).

Ana Moreira  
Presidente do Comitê de Programa



# Prefacio

Con esta edición 2008 en Gijón (7 al 10 de Octubre), las Jornadas de Ingeniería del Software y Bases de Datos (JISBD) celebra 13 años de existencia. JISBD nació como un foro donde la comunidad española publicaba su trabajo, discutía potenciales colaboraciones en investigación y evaluaba el progreso de los proyectos de investigación financiados por el Ministerio de Ciencia y Tecnología, y en la actualidad ha traspasado fronteras y cruzado varios océanos.

Actualmente, la conferencia es una referencia importante para jóvenes investigadores, así como un foro de cita obligada para investigadores más experimentados. Durante los últimos años, JISBD se ha abierto al mundo, aceptando artículos en Inglés y Portugués, además de Castellano. Este cambio no solamente se ha traducido en más participantes, sino que ha incrementado significativamente el número de artículos enviados, y principalmente, la calidad de los artículos aceptados.

La comunidad JISBD está actualmente auto sustentada y continúa expandiéndose. La calidad de los trabajos aceptados es equivalente al de otros eventos internacionales relevantes. Durante los últimos años, ha sido posible editar un volumen especial de IEEE LA con versiones ampliadas de los mejores trabajos presentados en la conferencia, lo que sucederá también en la presente edición. Este volumen especial, junto con las actas de la conferencia con ISBN, es una muestra de la calidad de los trabajos de JISBD.

Una de las características más sobresalientes de la conferencia ha sido la calidad de los ponentes invitados. Varios investigadores y profesionales de reconocido prestigio internacional han sido invitados a participar como ponentes en JISBD.

Dentro de este marco científico y tecnológico, la conferencia incluye varios eventos relacionados. Además de la presentación de artículos originales de alta calidad en la conferencia principal, el programa incluye tutoriales, demostraciones de herramientas, talleres para la discusión de ideas innovadoras y trabajos en curso, así como la divulgación de trabajos de investigación publicados en revistas y conferencias de prestigio (con un ratio de aceptación por debajo del 25% y un factor de impacto por encima de 0,5).

No es una exageración afirmar que JISBD ha consolidado su posición como un evento de referencia donde investigadores y profesionales de la Ingeniería del Software y las Bases de Datos se reúnen para discutir resultados y compartir ideas. JISBD se ha convertido en un foro importante para la colaboración entre diferentes sectores y grupos de investigación, en un contexto de excelente organización y excepcional hospitalidad.

## Sobre la presente edición

El crecimiento e internacionalización de JISBD se hace evidente analizando el origen de los artículos recibidos. En la presente edición, además de los artículos recibidos de los dos países de la Península Ibérica y los diez países Latinoamericanos, se han recibido artículos de China, Francia, Alemania, India, Irán y Pakistán.

De un total de 115 resúmenes previamente recibidos, finalmente se recibieron 112 artículos para su revisión. La mayoría de los artículos fueron revisados por tres miembros del Comité de Programa y varios por cuatro. El Comité de Programa aceptó 30 artículos largos y seleccionó 12 para su presentación como artículos cortos. El ratio de aceptación para los artículos largos fue de aproximadamente el 25%.

El éxito de la conferencia implica grandes responsabilidades en términos de garantizar la independencia de las revisiones y el cumplimiento de los estándares internacionales de ética. Por esta razón, durante los últimos años se ha realizado un mayor esfuerzo en aras de evitar envíos duplicados, una tarea especialmente dificultosa, ya que la conferencia acepta envíos en tres idiomas. En la

presente edición tres artículos fueron rechazados debido al doble envío en diferentes idiomas para diferentes eventos.

Además de los artículos aceptados, la conferencia incluye cinco talleres, un tutorial, nueve demostraciones de herramientas y foro para la discusión y divulgación de trabajos relevantes previamente publicados, así como una mesa redonda de carácter industrial.

Una característica importante de la conferencia es, sin ninguna duda, la excelencia de los ponentes invitados. La presente edición no es una excepción y estamos orgullosos de contar con la presencia de Bashar Nuseibeh y Bran Selic.

Bashar Nuseibeh es académico e investigador en la Open University del Reino Unido y profesor invitado en otras muchas universidades, incluyendo el Instituto Nacional Japonés de Informática. Bashar preside varios comités internacionales y está reconocido igualmente por su trabajo industrial, incluyendo organizaciones tales como el Servicio Nacional de Tráfico Aéreo del Reino Unido (NATS), Texas Instruments, Praxis Critical Systems, Philips Research Labs y la NASA.

Bran Selic fué durante varios años un destacado ingeniero e investigador en IBM y actualmente lidera una consultora internacional con sede en Canadá. Es internacionalmente conocido por su trabajo en sistemas industriales a gran escala y por su trabajo pionero en Desarrollo Dirigido por Modelos y Sistemas Empotrados en Tiempo Real.

La conferencia de Bashar se titula *“The five W’s (and one “H”) of Security: ... Software Engineering of Secure Systems”* y la de Bran *“Model-Based Software Engineering: Expected and Unexpected Challenges”*.

## Agradecimientos

Un agradecimiento especial es para Bashar y Bran por haber aceptado mi invitación y por compartir con todos los participantes sus conocimientos, experiencia y refinado sentido del humor.

Agradecimientos también para la multitud de colaboradores sin los cuales el éxito de la conferencia no habría sido posible. En primer lugar, para los autores de los artículos por confiar en la calidad de JISBD y enviar sus trabajos. En segundo lugar, para los miembros del Comité de Programa, cuyas revisiones aseguran la calidad de los trabajos.

Para el proceso de gestión y revisión de los trabajos recibidos, fui afortunada por tener la ayuda constante de Juan Hernández y José Javier Berrocal. Ellos fueron mis ángeles guardianes, alertándome constantemente de las fechas límite y siempre preparados para ayudarme cuando lo necesitaba. Agradecimientos especiales por la contribución de mi “Comité de Programa Ejecutivo”, Antonio Vallecillo, Juan Hernández, Miguel Toro, Vicente Pelechano y Xavier Franch.

Agradecimientos también para los organizadores de la conferencia principal, especialmente al presidente del comité organizador Javier Tuya y su vicepresidente Claudio de la Riva, por su manejo eficiente de las numerosas tareas que una conferencia de este tamaño y calidad conllevan. Agradecimientos también para los responsables de los eventos relacionados (en orden alfabético) António Rito Silva, Antonio Vallecillo, Gustavo Rossi, João Araújo, João Falcão e Cunha, José Berrocal, José Corrales, José García-Fanjul, João Miguel Fernandes, Lidia Fuentes y María José Suárez-Cabal.

Finalmente, palabras especiales de agradecimiento para los patrocinadores de la conferencia, sin cuya contribución el evento habría sido menos encantador (y con una gastronomía menos refinada).

Ana Moreira  
Presidenta del Comité de Programa

# Conference Committee

## **Program Committee Chair**

Ana Moreira (Univ. Nova de Lisboa, Portugal)

## **Organizing Chair**

Javier Tuya (Univ. Oviedo, Spain)

## **Organizing Co-Chair**

Claudio de la Riva (Univ. Oviedo, Spain)

## **Permanent Committee Secretary**

Mario Piattini (Univ. Castilla-La Mancha, Spain)

## **Tutorial Chair**

António Rito Silva (Univ. Técnica Lisboa, Portugal)

## **Workshop Chair**

João Araújo (Univ. Nova de Lisboa, Portugal)

## **Tool Demonstrations Chair**

Lidia Fuentes (Univ. Málaga, Spain)

## **Relevant Papers Dissemination Chairs**

Antonio Vallecillo (Univ. Málaga, Spain)

João Falcão Cunha (Univ. Porto, Portugal)

## **Proceedings Chair**

María José Suárez-Cabal (Univ. Oviedo, Spain)

## **Cyber Chair**

Jose Javier Berrocal (Univ. Extremadura, Spain)

## **Web Chair**

José A. Corrales (Univ. Oviedo, Spain)

## **Publicity Chairs**

Gustavo Rossi (Univ. La Plata, Argentina)

José García-Fanjul (Univ. Oviedo, Spain)

João Miguel Fernandes (Univ. Minho, Portugal)

## **Organizing Committee (Univ. Oviedo, Spain)**

Javier Tuya  
Claudio de la Riva  
José García-Fanjul  
Isabel Sevilla  
María José Suárez-Cabal  
José Ramón de Diego  
Raquel Blanco  
Eugenia Díaz Fernández  
José A. Corrales  
Marta Fernández de Arriba

## **SISTEDES Executive Board**

### **President**

Miguel Toro (Univ. Sevilla, Spain)

### **Vice President**

Juan José Moreno (Univ. Polit. Madrid, Spain)

### **Secretary**

Nieves R. Brisaboa (Univ. Coruña, Spain)

### **Treasurer**

Javier Tuya (Univ. Oviedo, Spain)

### **Members**

Pere Botella (Univ. Polit. Catalunya, Spain)  
Ricardo Peña (Univ. Complutense Madrid, Spain)  
Coral Calero (Univ. Castilla-La Mancha, Spain)  
Manuel Hermenegildo (Univ. Polit. Madrid, Spain)  
Ernesto Pimentel (Univ. Málaga, Spain)  
María Ribera Sancho (Univ. Polit. Catalunya, Spain)  
Natalia Juristo (Univ. Polit. Madrid, Spain)  
Salvador Lucas (Univ. Polit. Valencia, Spain)

# **Submission and Review Support System (Quercus Software Engineering Group)**

Javier Berrocal (Univ. Extremadura, Spain)  
Juan Hernández (Univ. Extremadura, Spain)

## **Secretariat**

Fundación Universidad de Oviedo  
C/ Principado 3, 4ª Planta  
33007 Oviedo, Spain.  
Tel: 34-985104927  
Fax: 34-985104928

## **Executive Program Committee**

Xavier Franch (Univ. Polit. Catalunya, Spain)  
Juan Hernández (Univ. Extremadura, Spain)  
Vicente Pelechano (Univ. Polit. Valencia, Spain)  
Antonio Vallecillo (Univ. Málaga, Spain)  
Miguel Toro (Univ. Sevilla, Spain)  
Javier Tuya (Univ. Oviedo, Spain)

## Program Committee

Albert Abelló (Univ. Polit. Catalunya, Spain)  
Ana Paula Afonso (Univ. Lisboa, Portugal)  
Ademar Aguiar (Univ. Porto, Portugal)  
Jesús Aguilar (Univ. Sevilla, Spain)  
José Aldana (Univ. Málaga, Spain)  
Mauricio Alférez (U. Nova de Lisboa, Portugal)  
Bárbara Álvarez (Univ. Polit. Cartagena, Spain)  
Raquel Anaya (Univ. EAFIT, Colombia)  
María José Aramburu (Univ. Jaume I, Spain)  
Hernán Astudillo (U. T. Federico Santa María, Chile)  
Orlando Belo (Univ. do Minho, Portugal)  
Rafael Berlanga (Univ. Jaume I, Spain)  
Paulo Borba (Univ. Federal Pernambuco, Brazil)  
Pere Botella (Univ. Polit. Catalunya, Spain)  
Rosana Braga (Univ. São Paulo, Brazil)  
Nieves Brisaboa (Univ. Coruña, Spain)  
Isabel Brito (Inst. Polit. Beja, Portugal)  
Fernando Brito e Abreu (U. Nova de Lisboa, Portugal)  
Coral Calero (Univ. Castilla-La Mancha, Spain)  
Marcelo Campo (UNICEN, Argentina)  
Carlos Canal (Univ. Málaga, Spain)  
Valeria de Castro (Univ. Rey Juan Carlos, Spain)  
Matilde Celma (Univ. Polit. Valencia, Spain)  
Christina Chávez (Univ. Bahia, Brazil)  
Rafael Corchuelo (Univ. Sevilla, Spain)  
Dolors Costal (Univ. Polit. Catalunya, Spain)  
Yania Crespo (Univ. Valladolid, Spain)  
Carlos Delgado (Univ. Carlos III, Spain)  
Oscar Díaz (Univ. País Vasco, Spain)  
Javier Dolado (Univ. País Vasco, Spain)  
Xavier Franch (Univ. Polit. Catalunya, Spain)  
Pablo de la Fuente (Univ. Valladolid, Spain)  
Mario Gaspar da Silva (Univ. Lisboa, Portugal)  
Alessandro García (Univ. Lancaster, UK)  
Marcela Genero (Univ. Castilla-La Mancha, Spain)  
Cristina Gómez (Univ. Polit. Catalunya, Spain)  
Jaime Gómez (Univ. Alicante, Spain)  
Alfredo Goñi (Univ. País Vasco, Spain)  
Silvia Gordillo (UNLP, Argentina)  
Pedro Guerreiro (Univ. Algarbe, Portugal)  
Juan Hernández (Univ. Extremadura, Spain)  
Jon Iturrioz (Univ. País Vasco, Spain)  
Elena Jurado (Univ. Extremadura, Spain)  
Natalia Juristo (Univ. Polit. Madrid, Spain)  
Miguel Katrib (Grupo WEBOO, Cuba)  
María Lencastre (Univ. Pernambuco, Brazil)  
Antonia Lopes (Univ. Lisboa, Portugal)  
Adolfo Lozano (Univ. Extremadura, Spain)  
Esperanza Marcos (Univ. Rey Juan Carlos, Spain)  
Henrique Madeira (Univ. Coimbra, Portugal)  
Eduardo Mena (Univ. Zaragoza, Spain)  
Ana María Moreno (Univ. Polit. Madrid, Spain)  
Juan José Moreno (Univ. Polit. Madrid, Spain)  
Juan Manuel Murillo (Univ. Extremadura, Spain)  
Oscar Pastor (Univ. Polit. Valencia, Spain)  
Vicente Pelechano (Univ. Polit. Valencia, Spain)  
Marcelo Pimenta (Univ. F. Rio Grande do Sul, Brazil)  
Ernesto Pimentel (Univ. Málaga, Spain)  
Mónica Pinto (Univ. Málaga, Spain)  
Ángeles Places (Univ. Coruña, Spain)  
Antonio Polo (Univ. Extremadura, Spain)  
Claudia Pons (UNICEN, Argentina)  
Tom Price (Univ. F. Rio Grande do Sul, Brazil)  
Carme Quer (Univ. Polit. Catalunya, Spain)  
Celia Ramos (Univ. Algarbe, Portugal)  
Isabel Ramos (Univ. Sevilla, Spain)  
Isidro Ramos (Univ. Polit. Valencia, Spain)  
Claudio de la Riva (Univ. Oviedo, Spain)  
José Riquelme (Univ. Sevilla, Spain)  
José Luis Roda (Univ. La Laguna, Spain)  
María José Rodríguez Fortis (Univ. Granada, Spain)  
José Raúl Romero (Univ. Córdoba, Spain)  
Antonio Ruiz (Univ. Sevilla, Spain)  
Francisco Ruiz (Univ. Castilla-La Mancha, Spain)  
José Samos (Univ. Granada, Spain)  
Fernando Sánchez (Univ. Extremadura, Spain)  
Juan Sánchez (Univ. Polit. Valencia, Spain)  
Carla Silva (Univ. F. Pernambuco, Brazil)  
Ernest Teniente (Univ. Polit. Catalunya, Spain)  
Miguel Toro (Univ. Sevilla, Spain)  
Ambrosio Toval (Univ. Murcia, Spain)  
Juan Carlos Trujillo (Univ. Alicante, Spain)  
Toni Urpi (Univ. Polit. Catalunya, Spain)  
Antonio Vallecillo (Univ. Málaga, Spain)  
Belén Vela (Univ. Rey Juan Carlos, Spain)

## Referees

Álvaro E. Prieto Ramos  
Amador Durán Toro  
André L. Santos  
Andrea Delgado  
Ángel Herranz  
Angélica Caro  
Anna Grimán Padua  
Antonio Jesús Roa Valverde  
Antônio Oliveira Filho  
Antonio Ruiz-Cortés  
Arturo Zambrano  
Carlos Bobed  
Carlos D. Barranco González  
Carlos Enrique Cuesta Quintero  
Carlos Neil  
Cecilia Delgado Negrete  
César J. Acuña  
Claudio Sant' Anna  
Cristina Vicente Chicote  
Daniel Rodríguez  
Dante Carrizo  
Diana Marcela Sánchez  
Diego Alonso Cáceres  
Diego Seco Naveiras  
Domingo Savio Rodríguez Baena  
Eduardo Rodríguez López  
Elisa Yumi Nakagawa  
Ellen Francine Barbosa  
Encarna Sosa Sánchez  
Fernando Molina Molina  
Fran J. Ruiz Bertol  
Francisco Javier Lucas Martínez  
Francisco Luís Gutiérrez Vela  
Francisco Martínez Álvarez  
Ignacio García Rodríguez de Guzmán  
Ismael Caballero  
Ismael Navas Delgado  
Ismael Sanz Blasco  
Javier Pérez García  
Joaquín Lasheras  
Joaquín Nicolás  
Jorge Gracia  
Jorge Martínez Gil  
José María Cavero Barca  
Juan Ángel Pastor Franco  
Juan M. Vara  
Juan Manuel Pérez Martínez  
Manuel Ángel Serrano Martín  
Manuel Resinas  
Márcio de Medeiros Ribeiro  
Marcirio Chaves  
Marcos López Sanz  
Mari Carmen Otero  
María Esperanza Manso Martínez  
María Luisa Rodríguez Almendros  
María Teresa Gómez López  
María Visitación Hurtado Torres  
Martin Solari  
Miguel Ángel Laguna Serrano  
Miguel Ángel Martínez  
Miguel Rodríguez Luaces  
M<sup>a</sup> Ángeles Moraga de la Rubia  
Nuno Cardoso  
Orlando Avila-García  
Oscar Dieste  
Óscar Pedreira Fernández  
Othmane Chniber  
Pablo Inostroza  
Pablo Trinidad  
Paloma Cáceres García de Marina  
Pedro Sánchez Palma  
Raquel M. Crespo García  
Raquel Trillo Lado  
Roberto Almeida Bittencourt  
Roberto Rodríguez Echeverría  
Roberto Ruiz  
Rui Lopes  
Sascha Ossowski  
Sergio Ilarri Artigas  
Vicente Luque Centeno

## Sponsors



Ayuntamiento de Gijón



GOBIERNO DEL  
PRINCIPADO DE ASTURIAS



INTERSYSTEMS





# Table of Contents<sup>1</sup>

## Keynote Address 1

|  |   |
|--|---|
| The five W's (and one "H") of Security: Software Engineering of Secure Systems ..... | 1 |
| <i>Bashar Nuseibeh</i>   |   |

## Aspects

|   |   |
|---|---|
| Analysis of Modularity by an Aspect-Oriented Measurement Process..... | 3 |
| <i>José Conejero, Juan Hernández, Elena Jurado, Klaas Berg</i>        |   |

## Process Engineering

|   |    |
|---|----|
| Automating the Software Process Management.....                 | 15 |
| <i>Javier Berrocal, José Manuel García, Juan Manuel Murillo</i> |    |

## Software Product Lines

|   |    |
|---|----|
| Generación Automática de Casos de Prueba en Líneas de Producto .....                  | 27 |
| <i>Pedro Mateo, Beatriz Lamanha, Macario Usaola</i>                                   |    |
| Gestión de la Variabilidad de los Requisitos de Seguridad en Líneas de Producto ..... | 39 |
| <i>Daniel Mellado, Eduardo Fernandez-Medina, Mario Piattini</i>                       |    |

---

<sup>1</sup> The section headings below correspond to the conference program, but do not include all the presentations in each conference session (where short papers and dissemination papers on the same topic also were included). Thus, the sections here all contain fewer papers than the corresponding conference session; the short papers are listed separated in this volume, followed by a chapter with an overview of the dissemination papers.

## Information Engineering

|  |    |
|--|----|
| Clasificación de Imágenes en el Sistema Qatris Imanager Mediante Regresión Logística Bayesiana ..... | 51 |
| <i>Inés Horrillo, Manuel Barrena</i>   |    |
| Efficient Retrieval of Ontology Fragments Using an Interval Labeling Écheme ...                      | 63 |
| <i>Victoria Romero, Rafael Llavori</i>   |    |
| Un Modelo para el Análisis y Explotación de Información Cognitiva en Repositorios Documentales ..... | 75 |
| <i>Miguel A. Martínez-Prieto, Joaquín Adiego, Pablo de la Fuente</i>                                 |    |
| Un Sistema de Consulta sobre Documentos Transformados con LZCS.....                                  | 87 |
| <i>Joaquín Adiego, Gonzalo Navarro, Pablo de la Fuente</i>   |    |

## Model Engineering

|   |     |
|---|-----|
| Análisis de Series Temporales Dirigido por Modelos Conceptuales sobre Datos Multidimensionales..... | 99  |
| <i>Jose Zubcoff, Jesús Pardillo, Juan Trujillo</i>  |     |
| Una Aproximación Dirigida por Modelos para el Desarrollo de Esquemas XML.....                       | 111 |
| <i>Verónica Bollati, Juan Vara, Belén Vela, Esperanza Marcos</i>                                    |     |
| Generación de Metadatos OLAP Dirigida por Modelos sobre Almacenes de Datos .....                    | 123 |
| <i>Juan Trujillo, Jesús Pardillo, Jose-Norberto Mazón</i>   |     |

## Formal Methods

|   |     |
|---|-----|
| Modelling Mash-up Resources .....   | 135 |
| <i>Iván Pérez, Ángel Herranz, Susana Muñoz, Juan Moreno-Navarro</i>               |     |
| Optimizando el Funcionamiento del Algoritmo FOIL .....                            | 147 |
| <i>Pablo Palacios, José Arjona, José Álvarez, Iñaki Fernández de Viana</i>        |     |
| Towards the Correctness Verification of Business Processes Modelled with UML..... | 159 |
| <i>Luis Mendoza, Manuel Capel, Kawtar Akhlaki</i>                                 |     |

## **Maintenance and Testing**

|   |     |
|---|-----|
| Agil_MANTEMA: Una Metodología de Mantenimiento de Software para Pequeñas Organizaciones ..... | 171 |
| <i>Francisco Pino, Francisco Ruiz, Jorge Triñanes, Félix García, Mario Piattini</i>           |     |
| Priorización del Valor de Artefactos Software Basada en la Frecuencia de Uso..                | 183 |
| <i>Daniel Cabrero, Javier Garzas, Mario Piattini</i>  |     |
| Identificación de Fallos en Módulos Software .....  | 195 |
| <i>José Riquelme, Roberto Ruiz, Daniel Rodríguez</i>  |     |

## **Data Mining, Data Streaming and Datawarehouses**

|   |     |
|---|-----|
| Hacia la Implementación Automática de Almacenes de Datos Seguros en Herramientas OLAP.....  | 205 |
| <i>Carlos Blanco, Ignacio García-Rodríguez de Guzmán, Eduardo Fernández-Medina, Juan Trujillo, Mario Piattini</i>                 |     |
| Una aproximación Basada en Diagramas de Actividades de UML para el Modelado Conceptual de Procesos ETL en Almacenes de Datos..... | 217 |
| <i>Lilia Muñoz, Jose-Norberto Mazón, Jesús Pardillo, Juan Trujillo</i>  |     |
| MeCADI*: un Marco Orientado a Objetivos para el Modelado de la Calidad en Almacenes de Datos.....                                 | 229 |
| <i>Cristina Cachero, Jesús Pardillo, Jose-Norberto Mazón, Juan Trujillo</i>   |     |

## **Reengineering and Software Modernization**

|  |     |
|--|-----|
| Reverse Engineering of Object-Relational Database Schemas .....      | 241 |
| <i>Jordi Cabot, Cristina Gómez, Elena Planas, M. Elena Rodríguez</i> |     |

## **Quality, Measurement & Estimation of Products & Processes**

|  |     |
|--|-----|
| Una Metodología Basada en ISO/IEC 15939 para la Elaboración de Planes de Medición de Calidad de Datos..... | 253 |
| <i>Eugenio Verbo, Ismael Caballero, Ricardo Pérez, Coral Calero, Mario Piattini</i>                        |     |
| Metodologías para Definir Programas de Medición en PyMEs: El Marco MIS-PyME.....                           | 265 |
| <i>María Díaz-Ley, Félix García, Mario Piattini</i>  |     |

|  |     |
|--|-----|
| Visualización de la Usabilidad de Componentes Software.....  | 275 |
| <i>M<sup>a</sup> Ángeles Moraga, Sergio Susín, Virginia Arcos, Coral Calero</i>                            |     |
| Aportaciones de una Visualización Metafórica al Análisis de Proyectos Software .....                       | 287 |
| <i>Amaia Aguirregoitia, J.Javier Dolado</i>  |     |
| Aplicación de las Técnicas de Modelado y Simulación en la Gestión de la Capacidad de los Servicios TI..... | 299 |
| <i>Elena Orta Cuevas, Mercedes Ruiz Carreira, Miguel Toro Bonilla</i>                                      |     |
| Measure Assessment for Heterogeneous XML Collections.....  | 311 |
| <i>María Pérez Catalán, Ismael Sanz, Rafael Berlanga</i>   |     |

## **Requirements Engineering**

|  |     |
|--|-----|
| Revisiones Sistemáticas: Recomendaciones para un Proceso Adecuado a la Ingeniería del Software .....           | 321 |
| <i>Oscar Dieste, Anna Grimán, Marta López</i>  |     |
| Metodologías Ágiles desde la Perspectiva de la Especificación de Requisitos Funcionales y No-Funcionales ..... | 333 |
| <i>Pilar Rodríguez, Agustín Yagüe, Pedro Alarcón, Juan Garbajosa</i>   |     |
| Metamodelo y Perfil UML para el Modelado Orientado a Metas de Requisitos Medibles.....                         | 345 |
| <i>Fernando Molina, Cristina Cachero, Jesús Pardillo, Ambrosio Toval</i>                                       |     |

## **Keynote Address 2**

|   |     |
|---|-----|
| Model-Based Software Engineering: Expected and Unexpected Challenges..... | 357 |
| <i>Bran Selic</i>   |     |

## **Short Papers**

|  |     |
|--|-----|
| AAJ: Un Lenguaje de Descripción Arquitectónica Orientado a Aspectos..... | 361 |
| <i>María Boton, Amparo Navasa</i>  |     |
| An Ontology for IT Services .....  | 367 |
| <i>Jorge Freitas, Anacleto Correia, Fernando Abreu</i>                   |     |

|  |     |
|--|-----|
| Construcción de Modelos Lógicos Multidimensionales Seguros<br>para su Implementación en Herramientas OLAP Mediante MDA y QVT .....               | 373 |
| <i>Carlos Blanco, Ignacio García-Rodríguez de Guzmán,<br/>Eduardo Fernández-Medina, Juan Trujillo, Mario Piattini</i>                            |     |
| Desarrollo de Almacenes de Datos Espacio Temporales Dirigido por Modelos ..  | 379 |
| <i>Octavio Glorio, Juan Trujillo</i>   |     |
| Generating Domain Specific Aspect Code for Navigation from Platform<br>Specific Models in MWACSL.....  | 385 |
| <i>Antonia M. Reina Quintero, Miguel Toro Bonilla, Jesús Torres Valderrama</i>   |     |
| Zentipede: Una Contribución a la Renovación de la Gestión del Proceso<br>Software .....  | 391 |
| <i>José Manuel García Alonso, José Javier Berrocal,<br/>Juan Manuel Murillo Rodríguez</i>  |     |
| Hacia la Definición de un Simulador para la Enseñanza de la Elicitación<br>de Requisitos en el Contexto del Desarrollo Global del Software ..... | 417 |
| <i>Miguel Romero, Aurora Vizcaino, Mario Piattini</i>  |     |
| Un Marco de Referencia para Comparar ESBs desde la Perspectiva<br>de la Integración de Aplicaciones.....   | 403 |
| <i>Rafael Corchuelo, Rafael Frantz, Jesús González</i>   |     |
| Refactorizaciones en la Migración del Software.....  | 409 |
| <i>Rául Marticorena, Yania Crespo, Carlos López</i>  |     |
| Diseño Evolutivo de Bases de Datos XML .....   | 415 |
| <i>Carlos Nilo, Cecilia Reyes, Jose Marti</i>  |     |
| Impacto de las Multiplicidades en la Resolución de Problemas<br>de Sumarizabilidad para OLAP .....   | 421 |
| <i>Jose-Norberto Mazón, Jens Lechtenbörger, Juan Trujillo</i>  |     |

## **Workshops, Tutorials, Demos and Dissemination**

|                           |            |
|---------------------------|------------|
| <b>Workshops.....</b>     | <b>427</b> |
| <i>João Araújo</i>        |            |
| Tutorials .....           | 429        |
| <i>António Rito Silva</i> |            |

|  |            |
|--|------------|
| <b>Tool Demonstrations .....</b>   | <b>431</b> |
| <i>Lidia Fuentes</i>   |            |
| ActiveRulesDBX – Ferramenta para Execução de Regras a partir da Detecção de Eventos Temporais.....   | 433        |
| <i>Eugênio de O. Simonetto, Jéferson Kasper, Giovanni R. Librelotto</i>  |            |
| Deriving AO Software Architectures using the AO-ADL Tool Suite .....   | 437        |
| <i>Mónica Pinto, Lidia Fuentes, Luis Fernández, Juan A. Valenzuela</i>   |            |
| ESFORA: a tool for the dEfinition of domain SPecific OpeRation languages.....  | 441        |
| <i>David Musat, Jennifer Pérez, Pedro P. Alarcón, Agustín Yagüe</i>  |            |
| FAMA Framework .....   | 445        |
| <i>Pablo Trinidad, David Benavides, Antonio Ruiz-Cortés, Sergio Segura</i>   |            |
| ProSÉ: A Protégé plugin for Reusing Ontologies, Safe and Économique .....  | 449        |
| <i>Ernesto Jiménez-Ruiz, Bernardo Cuenca Grau, Ulrike Sattler<br/>Thomas Schneider, Rafael Berlanga</i>                                    |            |
| REMM-Studio+: Extensiones para Modelar Variabilidad y Permitir la Reutilización de Requisitos .....  | 453        |
| <i>Begoña Moros, Cristina Vicente-Chicote, Ambrosio Toval</i>  |            |
| RUX-Tool: Una herramienta CASE para el modelado y la generación automática de Interfaces de Usuario para RIA .....                         | 457        |
| <i>Marino Linaje, Juan Carlos Preciado, Fernando Sánchez-Figueroa<br/>Rober Morales-Chaparro, David Gordillo, Fernando Sánchez-Herrera</i> |            |
| StateML+: Diseño, Validación y Generación de Código Ada para Máquinas de Estado Jerárquicas .....  | 461        |
| <i>Diego Alonso, Cristina Vicente-Chicote, Bárbara Álvarez</i>   |            |
| <b>Relevant Papers Dissemination .....</b>   | <b>465</b> |
| <i>Antonio Vallecillo, João Falcão Cunha</i>   |            |
| Feature Oriented Model Driven Development: A Case Study for Portlets.....  | 467        |
| <i>Salvador Trujillo, Don Batory, Oscar Díaz</i>   |            |
| DEX: High-Performance Exploration on Large Graphs for Information Retrieval.....   | 69         |
| <i>Norbert Martínez-Bazan, Victor Muntés-Mulero, Sergio Gómez-Villamor,<br/>Jordi Nin, Mario-A. Sánchez-Martínez, Josep-L. Larriba-Pey</i> |            |
| Determining Criteria for Selecting Software Components: Lessons Learned .....  | 471        |
| <i>Juan Pablo Carvallo, Xavier Franch, Carme Quer</i>  |            |

|   |            |
|---|------------|
| Engineering Rich Internet Application User Interfaces over Legacy Web Models .....                      | 473        |
| <i>Marino Linaje, Juan Carlos Preciado, Fernando Sánchez-Figueroa</i>                                   |            |
| Guideliness for Eliciting Usability Functionalities .....   | 475        |
| <i>Natalia Juristo, Ana María Moreno, María-Isabel Sánchez-Segura</i>                                   |            |
| From Wrapping to Knowledge .....  | 477        |
| <i>José Luis Arjona, Rafael Corchuelo, David Ruiz, Miguel Toro</i>                                      |            |
| Introducing Structure Management in Automatic Reference Resolution: An XML-based Approach .....         | 479        |
| <i>M. Mercedes Martínez-González, Pablo de la Fuente</i>  |            |
| Run-time Composition and Adaptation of Mismatching Behavioural Transactions .....                       | 481        |
| <i>Javier Cámara, Gwen Salaün, Carlos Canal</i>   |            |
| Building Domain-Specific Languages for Model-Driven Development .....                                   | 483        |
| <i>Jesús Sánchez Cuadrado, Jesús García Molina</i>  |            |
| Reconciling requirement-driven data warehouses with data sources via multidimensional normal forms..... | 485        |
| <i>Jose-Norberto Mazón, Juan Trujillo, Jens Lechtenbörger</i>   |            |
| Developing Secure Data Warehouses with a UML Extension.....   | 487        |
| <i>Eduardo Fernández-Medina, Juan Trujillo, Rodolfo Villarroel, Mario Piattini</i>                      |            |
| <b>Author Index.....</b>  | <b>489</b> |

# Una Metodología Basada en ISO/IEC 15939 para la Elaboración de Planes de Medición de Calidad de Datos

Eugenio Verbo<sup>1</sup>, Ismael Caballero<sup>1,2</sup>, Ricardo Pérez<sup>1</sup>, Coral Calero<sup>2</sup>, Mario Piattini<sup>2</sup>

<sup>1</sup> Indra Software Labs, Departamento de I+D de Indra Software Labs,  
Ronda de Toledo s/n 13004 Ciudad Real, España  
{emverbo, icaballerom, rpdcastillo}@indra.es

<sup>2</sup> Grupo de Investigación ALARCOS, Instituto de Tecnologías de la Información y Sistemas,  
Paseo de la Universidad 4, 13071, España  
{Ismael.Caballero, Coral.Calero, Mario.Piattini}@uclm.es

**Resumen.** Hoy en día, los datos juegan un papel fundamental en las organizaciones y la gestión de su calidad se está convirtiendo en una actividad imprescindible. Como parte de dicha gestión, y en aras de obtener medidas útiles, las organizaciones necesitan realizar planes de medición de Calidad de los Datos (CD). Estos planes de medición deben hacerse teniendo en cuenta la propia naturaleza de los datos así como los factores organizacionales que afectan al uso de los mismos. Dado que no existen muchos trabajos en la bibliografía con este objetivo, este artículo presenta una metodología, MEPLAMECAL, para elaborar planes de medición de CD. MEPLAMECAL está basada en ISO/IEC 15939 que, a pesar de ser un estándar de calidad de software, consideramos que puede ser aplicado en este contexto debido a las similitudes existentes entre software y datos. La metodología propuesta está compuesta por dos actividades: (1) establecer y mantener el compromiso de la organización en el proceso de medición, y (2) la elaboración del plan de medición. Estas dos actividades se estructuran en tareas para las que se proponen productos de entrada y salida, así como técnicas y herramientas a utilizar, muchas de ellas tomadas de la Ingeniería del Software.

**Palabras Clave:** Calidad de Datos, Medición de la Calidad de Datos, ISO/IEC 15939, Modelo de Información de Medición de la Calidad de Datos.

## 1 Introducción

Levis et al. en [26] mencionan ejemplos de escenarios en los que datos con niveles de calidad inadecuados originan problemas que afectan negativamente a los Sistemas de Información y, por tanto, al rendimiento organizacional. Las causas más comunes de esos niveles inadecuados son una serie de obstáculos a lo largo del ciclo de vida de los datos en el SI como los descritos por Strong et al. en [33]. Por su parte, los problemas producidos por esta falta de calidad pueden ser clasificados a diferentes niveles según su naturaleza: técnicos (como los referidos a la implementación de almacenes de datos [29]), organizacional (tales como pérdida de clientes [31], grandes pérdidas financieras [13, 27] o incluso insatisfacción de los trabajadores [12, 33]) y legales (como la violación del artículo cuarto del Apartado II de la LOPD de 1999).

Con el objetivo de minimizar el impacto negativo de estos problemas en el desarrollo de sus actividades, es fundamental que las organizaciones sean capaces de evaluar si el nivel de CD de sus datos es el adecuado. Esto implica por un lado la definición de medidas de CD sobre ellos y por otro el establecimiento de unos rangos de aceptación válidos para los valores obtenidos en esas medidas. De esta forma, y aplicando las técnicas y herramientas clásicas de calidad será posible localizar de forma más eficiente los datos con niveles de CD inadecuados y sus causas.



Sin embargo, debido a la propia naturaleza de los datos, puede ser bastante difícil realizar una definición adecuada de las medidas. Una aproximación que puede hacer más sencillo el proceso de definición de medidas, es la propuesta por el Programa TDQM del MIT en [32, 37], en la que se aporta una visión desde las teorías clásicas de calidad, en la que los datos pueden ser considerados como la materia prima de un proceso de producción en el que la información es el producto resultante. Teniendo en cuenta esta analogía de información con los clásicos productos manufacturados se habilita la aplicación de los principios clásicos de gestión de la calidad a la CD como algunos autores, de forma implícita o explícita, habían sugerido ([3, 6, 26]).

La bibliografía consultada describe algunas metodologías genéricas de evaluación de CD como la propuesta por Lee et al. en [25] (donde no se describen los procedimientos de medición como parte de la definición de las medidas), y otras particulares centradas en los problemas específicos de los escenarios donde se requería la medición como por ejemplo la propuesta por Al-Hakim en [1] para entornos médicos, donde habitualmente las medidas son realizadas *ad hoc*, sin una planificación previa y casi siempre sin tener en cuenta algunos aspectos importantes relativos tanto a la disponibilidad y disposición de los recursos organizacionales para el proceso de medición como a la naturaleza específica de los datos que afectan al propio proceso de medición. Desafortunadamente, esta suele ser una carencia común en los trabajos encontrados.

No obstante, algunos trabajos como los propuestos en [4, 17, 19, 28] analizan varios de los aspectos mencionados, y tratan de generalizar y priorizar sus conclusiones. Sin embargo, y a pesar de haber obtenido un conjunto de conceptos con bastantes elementos comunes, se da la circunstancia de que normalmente han usado diferentes términos para el mismo concepto, lo que dificulta la puesta en práctica de los resultados que se han obtenido en escenarios distintos de los suyos. Para tratar de paliar esta inconsistencia en la nomenclatura, Caballero et al. en [7] han analizado los términos de los autores más referenciados y han propuesto una terminología unificada de acuerdo con el Modelo de Información de Medición de Software propuesto en ISO/IEC 15939 [23] y la han ampliado con los aspectos específicos de CD. El resultado es un **Modelo de Información de Medición de Calidad de Datos** (*Data Quality Measurement Information Model*, **DQMIM**).

Un valor añadido de ISO/IEC 15939 es que proporciona una metodología para la definición de planes de medición de software observando aspectos organizacionales. Esta metodología podría ser adaptada y extendida con los aspectos específicos de CD a los que antes se hacía referencia para cubrir el hueco existente en el campo de CD. La principal contribución de este artículo al campo de la CD es **MEPLAMECAL**, una metodología basada en la propuesta por ISO/IEC 15939 para la elaboración de planes de medición de CD.

La definición de la metodología implica no sólo la identificación de las actividades y tareas correspondientes, sino también la identificación de los productos de entrada y salida para cada una de ellas. Además, se identifican, a partir de las utilizadas en el campo de Ingeniería del Software, algunas técnicas y herramientas útiles para transformar las entradas en salidas. Pensamos que esta estructura facilitará la incorporación de los aspectos de calidad de datos a las buenas prácticas habituales de calidad del software.

El resto del artículo se estructura de la siguiente manera: la sección 2 presenta un breve repaso al estado del arte de la medición de CD, incluyendo una sinopsis del DQMIM. La sección 3 describe la metodología en sí misma. La sección 4 perfila algunas conclusiones y el trabajo futuro; por último, los agradecimientos y referencias se pueden encontrar al final del artículo.

## 2 Factores que influyen en la Medición de la Calidad de Datos

Hay un gran número de definiciones para el concepto de CD [2]. De cualquier forma, la mayoría de los autores coinciden en que un dato es de calidad cuando es válido para el propósito para el que un usuario de esos datos quiere utilizarlo (“*adecuación al uso*” [2, 34]). De aquí en adelante, para ganar generalidad en las explicaciones, se ampliará el término “**usuario**” mediante el término

“**implicado**”, refiriéndose a *cualquier agente (persona, sistema o proceso) involucrado en el uso de los datos* [37].

La definición anterior de calidad, basada en la adecuación al uso, tiene dos componentes importantes: percepción multidimensional de la calidad y dependencia de contexto. Entorno a estos dos componentes, los implicados deben tener en cuenta los factores a los que se hacía referencia en la introducción de cara a la definición de las medidas de CD.

En cualquier caso, los implicados deben primero identificar las razones por las que quieren medir el nivel de CD de los datos usados por alguno de los procesos de la organización. De acuerdo al DQMIM, se puede decir que los implicados podrían querer satisfacer determinadas **Necesidades de Información**.

Una de las estrategias más usadas para tratar el estudio de la percepción multidimensional de la CD es descomponerla en características más pequeñas, igual que ISO/IEC 9126 [22] hace para el software. Estas características se denominan Dimensiones de CD [2, 24] aunque DQMIM propone el término “**concepto medible**” en su lugar.

La identificación y definición de estos conceptos medibles de CD es todavía un reto para la comunidad de CD debido a su dependencia del contexto. Una definición adecuada permitiría a los investigadores y expertos entender (y, por tanto, medir) la CD, evitando la ambigüedad en las definiciones, términos y conceptos utilizados. Algunos trabajos, como [2, 24, 36] son de obligada referencia para consultar el significado de los conceptos medibles de CD más usados. Para cada escenario en el que se necesita medir la CD de los datos usados, se deben elegir los conceptos medibles más adecuados. El conjunto de conceptos medibles de CD escogidos es conocido como **modelo de CD**. Existen numerosos ejemplos en las referencias bibliográficas de modelos de CD para distintos entornos: salud y asistencia sanitaria [1], militar [5], sistemas de apoyo a la decisión [18], o web [9, 14], por citar algunos. Es importante reseñar que ISO está trabajando actualmente en el estándar ISO/IEC 25012 [21], una parte de la familia de estándares SQUARE que propondrá un modelo de CD para SI. De cualquier forma, y a la espera de que el estándar esté terminado, la clasificación propuesta por [34] es la más utilizada (ver Tabla 1).

**Tabla 1.** Modelo genérico de CD por [34]

| Categoría           | Conceptos medibles de CD  | Descripción  |
|---------------------|---|--|
| CD intrínseca       | Precisión ( <i>Accuracy</i> ), Objetividad ( <i>Objectivity</i> ), Credibilidad ( <i>Believability</i> ), Reputación ( <i>Reputation</i> )  | Calidad que tienen los datos por sí mismos                             |
| CD de accesibilidad | Accesibilidad ( <i>Accessibility</i> ), Seguridad de Acceso ( <i>Access security</i> )  | Suministran significado sobre la facilidad de acceso a los datos       |
| CD contextual       | Relevancia ( <i>Relevancy</i> ), Valor añadido ( <i>Value-Added</i> ), Oportunidad ( <i>Timeliness</i> ), Completitud ( <i>Completeness</i> ), Cantidad de datos ( <i>Amount of data</i> )  | Tratan con el uso de los datos en un contexto                          |
| CD representacional | Interpretabilidad ( <i>Interpretability</i> ), Facilidad de comprensión ( <i>Ease of understanding</i> ), Representación concisa ( <i>Concise Representation</i> ), Representación consistente ( <i>Consistent representation</i> ) | Características de la representación de los datos que los hacen útiles |

Cualquier implicado, sea cual sea su rol, necesitará determinar el grado de bondad de un dato con respecto a los conceptos medibles más adecuados para la tarea que se esté realizando. Esta medida dependerá del uso previsto para los datos y de la naturaleza de los conceptos medibles de CD, que determinan el método o la función de medición [23]. Las típicas medidas de CD tienen una escala de ratio con valores suministrados por la función de medición como, por ejemplo, la fórmula propuesta por [2, 24] y que se muestra en (1):

$$CD_{Medida} = 1 - \frac{\text{NúmeroDeUnidadesDeDatosQueNoSatisfacenUnCriterio}}{\text{NúmeroTotalDeUnidadesDeDatos}} \quad (1)$$

En la fórmula (1), *UnidadesDeDatos* se refiere a las instancias de los atributos medibles. Además hay dos medidas base:

- *NúmeroTotalDeUnidadesDeDatos*, cuyos valores puede ser calculados de forma objetiva usando como método de medición el conteo de las instancias de los atributos medibles de las entidades cuya CD está siendo evaluada, y
- *NúmeroDeUnidadesDeDatosQueNoSatisfacenUnCriterio*, cuyos valores también pueden ser calculados de una forma objetiva contando el número de “falsos” obtenido después de comprobar si las unidades de datos satisfacen el criterio.

En algunas ocasiones, para asignar valores a la medición de un concepto medible y poder aplicar el criterio, es preciso completar mediante metadatos el significado de una unidad de datos de acuerdo a ese concepto medible. Para tomar una decisión, se necesita además de un valor para el metadato, una regla que usando los metadatos describa el criterio de aceptación como, por ejemplo, la pertenencia del valor del metadato a un dominio determinado. [28] identifica como fuentes de valores para metadatos a los propios implicados (se considera que normalmente proporcionaría un valor subjetivo), al proceso de producción de la información o incluso al mismo almacén de datos.

Para hacer repetibles los procesos de medición, es necesario que el valor del metadato quede adjuntado al dato que completa en el modelo de datos del SI. En [38], se propone una solución para el modelo relacional mediante el etiquetado de los datos como si fueran atributos relacionales. [7] propone un esquema XML llamado DQXSD que permite añadir etiquetas a ficheros XML, [8] incluso propone el uso de Tecnologías Semánticas para realizar esta adjuntado, habilitando así a las aplicaciones Web el procesamiento de los aspectos referidos a la medición de CD.

Otro factor importante que debe ser tenido en cuenta, es la posibilidad de limitar el número de unidades de datos en las que el criterio va a ser aplicado [12] de cara a minimizar el esfuerzo computacional invertido en la medición y que consumiría recursos no dedicados al procesamiento de los datos. Así, para mejorar el rendimiento del SI puede ser necesario seleccionar un conjunto representativo de unidades de datos. El número de unidades de datos y su distribución dependerá de la naturaleza de las necesidades de información.

Para no interferir en el proceso de medición, es necesario fijar en el ciclo de vida cuál es el mejor momento o el mejor intervalo de tiempo para ejecutar el proceso de medición sobre los datos [31].

Como se ha visto a lo largo de este apartado, gestionar todos estos factores puede resultar complicado. Por esto, comprendemos la necesidad de alguna guía para la definición de planes de medición de CD que tenga en cuenta estos factores y su influencia en el proceso de planificación de la medida. A continuación se presenta MEPLAMECAL.

### 3 Metodología MEPLAMECAL.

El objetivo de esta sección es describir de forma resumida los puntos más interesantes de MEPLAMECAL. La metodología se compone de dos actividades, descritas más adelante, divididas a su vez en tareas. Para cada una de las tareas propuestas, se enumeran tanto los principales productos de entrada como los productos de salida esperados. También se propone el uso de técnicas y herramientas (procedentes en su mayoría del campo de la Ingeniería del Software) para obtener los productos de salida, aunque la elección depende de las preferencias de cada organización. Con el objetivo de facilitar el uso de la metodología, también se identifican los implicados que deberían participar en la ejecución de cada tarea. Las actividades se etiquetan con COM y con EPM, y para las tareas se añade un número que indica el orden en el que deberían ser realizadas. En la Fig. 1 se muestra un diagrama que resume las actividades y tareas de la metodología, junto con su orden de ejecución y los implicados en cada una de ellas.

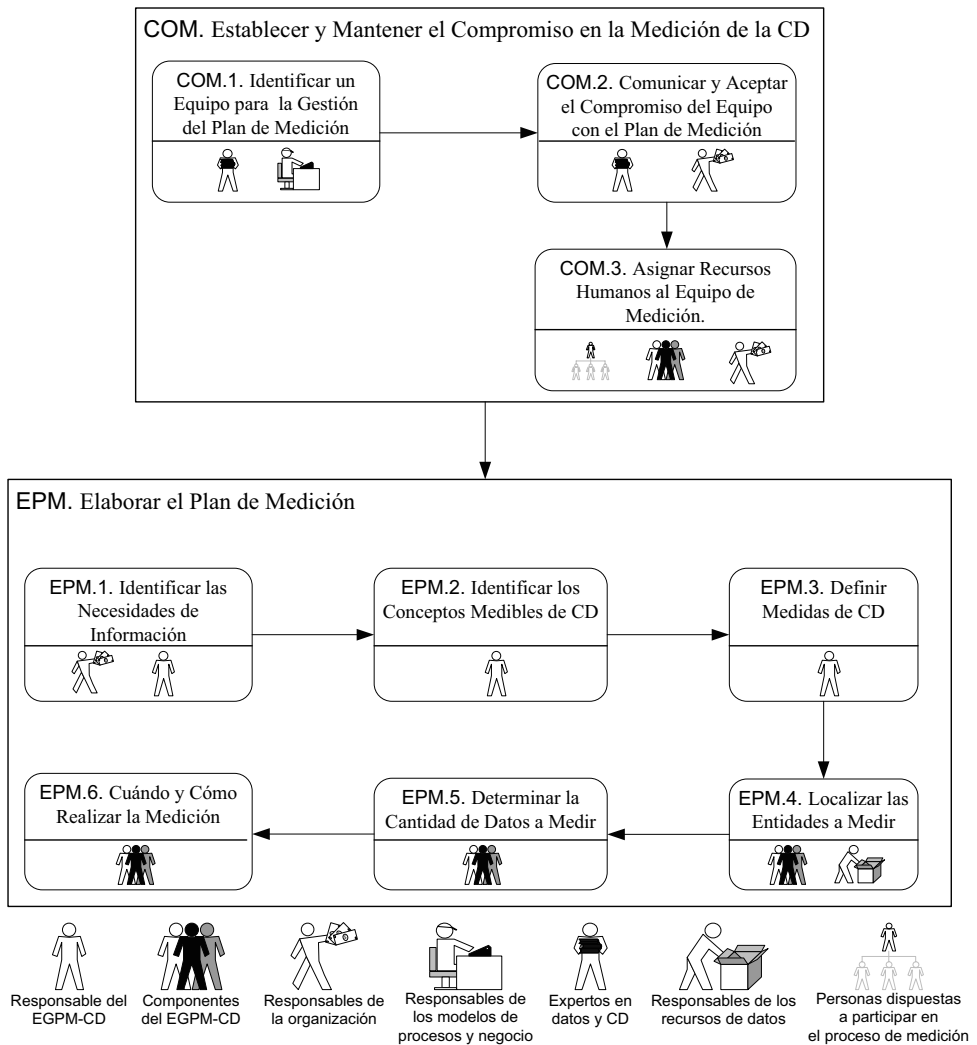


Fig. 1. Diagrama de la metodología MEPLAMECAL.

### 3.1 COM. Establecer y Mantener el Compromiso en la Medición de la CD.

Los expertos en calidad coinciden en la necesidad de delegar la responsabilidad de la gestión de la calidad a un grupo de gente comprometida de la organización [11]. Para dar soporte a esta necesidad, uno de los objetivos de la metodología es esbozar un equipo multidisciplinar de trabajadores, que puede responsabilizarse de la realización del plan de medición de la CD. Para satisfacer el objetivo propuesto en esta actividad, se deben ejecutar las siguientes tareas.

#### COM.1. Identificar un Equipo para la Gestión del Plan de Medición.

Este equipo será el responsable de coordinar los esfuerzos y recursos disponibles de la organización de cara a la planificación de la medición de la DQ. El **Equipo de Gestión del Plan de Medición de la CD (EGPM-CD)** debería estar compuesto por roles con responsabilidad directa sobre los datos y su integridad, de forma que se pueda determinar quién está usando los

datos y para qué propósito, con el fin de valorar la naturaleza y alcance de cualquier deficiencia que pueda existir, así como evaluar el impacto que los problemas relacionados con la CD puedan ocasionar. En consecuencia, si se decide externalizar el proceso de medición, el EGPM-CD no debería estar formado exclusivamente por personas externas a la organización sino que sería necesaria la cooperación entre componentes externos e internos para la elaboración del plan.

Dicho equipo debe ser multidisciplinar, cohesionado, sus miembros deben complementarse en conocimientos y cualidades, y tener capacidad de crítica a su propio trabajo. Además deben disponer de un entorno operativo donde se identifique un método de trabajo con formas efectivas de comunicación, así como contar con un catálogo de técnicas y herramientas útiles para cada tarea. La Tabla 2 muestra los principales artefactos para esta actividad.

**Tabla 2.** Artefactos para COM.1.

|                         |         |  |
|-------------------------|---------|--|
| Productos               | Entrada | - Lista de personas que gestionan o trabajan con los datos de la organización.                       |
|                         | Salida  | - Propuesta de Equipo de Gestión del Plan de medición (EGPM-CD).                                     |
| Herramientas y técnicas |         | - Entrevistas<br>- Sesiones de trabajo.  |
| Implicados              |         | - Expertos en datos y CD.<br>- Responsables de los modelos de procesos y negocio de la organización. |

### COM.2. Comunicar y Aceptar el Compromiso del Equipo con el Plan de Medición.

Una vez identificado el EGPM-CD, se debe llevar a cabo una segunda tarea de asignación de personas a los roles componentes del EGPM-CD. Sería interesante llevar a cabo entrevistas y sesiones de trabajo con los candidatos para comprobar su disposición, adecuación y, llegado el caso, su compromiso en la elaboración del plan de medición. La Tabla 3 da más detalles acerca de los artefactos de esta actividad.

**Tabla 3.** Artefactos para COM.2.

|                         |         |   |
|-------------------------|---------|---|
| Productos               | Entrada | - Propuesta de Equipo de Gestión del Plan de Medición (EGPM-CD).<br>- Organigrama de la organización.   |
|                         | Salida  | - Compromiso de los componentes de la organización para participar en la elaboración del plan de medición.<br>- Documentos de carga de trabajo. |
| Técnicas y herramientas |         | - Entrevistas.<br>- Sesiones de trabajo.<br>- Compromiso.   |
| Implicados              |         | - Expertos en datos y CD.<br>- Responsables de la organización.   |

### COM.3. Asignar Recursos Humanos al Equipo de Medición.

Tras conseguir el compromiso de los componentes del EGPM-CD propuesto, hay que estudiar su disponibilidad y restricciones temporales. Para ello, se deberían realizar entrevistas con los responsables de la organización para identificar posibles solapamientos con otros proyectos de la organización que pudieran malograr la ejecución del plan de medición. Algunas herramientas útiles para conseguir este objetivo son los diagramas de Gantt o PERT. Por otra parte, se deberían llevar a cabo sesiones de trabajo y entrevistas para detectar y clasificar habilidades individuales de los componentes del EGPM-CD y, dependiendo de ellas, asignar a cada participante las tareas en las que sería más útil. La Tabla 4 muestra los artefactos correspondientes.

**Tabla 4.** Artefactos para COM.3.

|                         |         |   |
|-------------------------|---------|---|
| Productos               | Entrada | - Componentes del EGPM-CD.  |
|                         | Salida  | - Composición definitiva del EGPM-CD.   |
| Técnicas y herramientas |         | - Sesiones de trabajo.<br>- Entrevistas.<br>- Herramientas de planificación temporal como los diagramas de Gantt o PERT.<br>- Estimaciones.<br>- Análisis de coste/beneficio. |
| Implicados              |         | - Responsables de la organización.<br>- Componentes del EGPM-CD.<br>- Personas dispuestas a participar en el proceso de medición.   |

### 3.2 EPM. Elaborar el Plan de Medición.

El principal objetivo de esta segunda actividad es esbozar el Plan de Medición de CD teniendo en cuenta todos los factores mencionados en la sección 2. Para documentar el plan se usarán los términos proporcionados por la terminología DQMIM. Como resultado se obtendrá un documento con el plan de medición de CD. Para conseguir este objetivo, se propone la realización de las siguientes tareas.

#### EPM.1. Identificar las Necesidades de Información.

Para cada uno de los escenarios donde sea necesaria la medición, es esencial recoger y comprender los requisitos de calidad que deben satisfacer los datos y comprobar que el funcionamiento actual del SI satisface dichos requisitos [17]. Si se encontraran no-conformidades (niveles inadecuados de CD), el EGPM-CD debe profundizar en las causas que las provocaron para arreglarlas o, al menos, mitigar sus efectos. La clasificación propuesta en [29] recoge causas comunes relacionados con los SI que pueden ser utilizados como guía en la identificación de dichas no-conformidades. Puesto que las necesidades de información no se refieren únicamente a aspectos técnicos sino también a organizativos y de gestión, se deberían tener en cuenta las relaciones de negocio existentes entre los principales usuarios del SI. La Tabla 5 muestra los artefactos para esta tarea.

**Tabla 5.** Artefactos para EPM.1.

|                         |         |  |
|-------------------------|---------|--|
| Productos               | Entrada | - Especificación de requisitos de CD.<br>- Informes de actividad de los SI de la organización. |
|                         | Salida  | - Lista de necesidades de información.   |
| Técnicas y herramientas |         | - Entrevistas.<br>- Sesiones de trabajo.   |
| Implicados              |         | - Responsables de EGPM-CD.<br>- Responsables de la organización.                               |

#### EPM.2. Identificar los Conceptos Medibles de CD.

Los conceptos medibles de CD son criterios racionales que representan los requisitos de usuario para juzgar la CD. El EGPM-CD se debe encargar de seleccionar aquellos conceptos medibles que mejor satisfagan las necesidades de información. Como base para la realización de esta tarea se pueden tomar de la bibliografía aquellos modelos de CD que mejor se adapten al escenario de medición o incluso, si estuvieran disponibles en la organización, aquellos elaborados a partir de experiencias previas. Sobre estos modelos se aplicarán técnicas como tormentas de ideas, sesiones de trabajo, entrevistas o el método Delphi. También se podría utilizar la metodología de Franch y Carvalho descrita en [15] pero adaptada a nuestro campo de calidad de datos. De todos modos, en nuestra metodología se propone continuar la estrategia propuesta por Strong et al. en [33]. En ella se identifican los obstáculos de CD más comunes y los conceptos medibles relativos a cada uno de

ellos. De esta forma, relacionando los obstáculos que afectan a las necesidades de información definidas anteriormente, se pueden obtener los conceptos medibles relativos a cada necesidad de información.

Finalmente, hay tener en cuenta que puede haber dependencias entre los conceptos medibles de CD que forman el modelo, tal y como [10, 16] han analizado, siendo necesario modelar cómo afectan al proceso de medición. En la Tabla 6, se resumen los principales artefactos para esta tarea.

**Tabla 6.** Artefactos para EPM.2.

|                         |         |   |
|-------------------------|---------|---|
| Productos               | Entrada | - Necesidades de información.<br>- Especificación de requisitos de CD.<br>- Catálogo de conceptos medibles de CD. |
|                         | Salida  | - Lista de conceptos medibles de CD relevantes.<br>- Lista de entidades a medir.                                  |
| Técnicas y herramientas |         | - Sesiones de trabajo.<br>- GQM.<br>- Catalogación.<br>- Lista de obstáculos de Strong et al. en [32]             |
| Implicados              |         | - Responsables del EGPM-CD.   |

### EPM.3. Definir Medidas de CD.

Una vez que los conceptos medibles de CD han sido identificados, es el momento de definir las medidas propiamente dichas, correspondientes a los atributos medibles de las entidades que contienen datos. Tres de las metodologías genéricas más importantes para definir medidas son IEEE 1061, ISO/IEC 15939 y “*Goal-Question-Metric*” (GQM).

De acuerdo al modelo propuesto en DQMIM, las medidas de CD pueden ser de uno de los siguientes tipos: medidas base, medidas derivadas o indicadores. Es importante resaltar que la definición de un plan de medición puede provocar cambios tanto en el modelo de procesos o modelo de datos del SI que contiene los datos cuya calidad se pretende medir. Un ejemplo de este caso es la inserción de metadatos tal y como se describió en la sección 2.

Para cada concepto medible de CD, se debe especificar un procedimiento de medición dependiendo del tipo de medida. Esto lleva aparejado la identificación de una unidad de medida, de una escala y de los correspondientes elementos para completar la definición de la medida. Por ejemplo, para un indicador se debe añadir un criterio de decisión y un modelo de análisis.

Dada la complejidad de la situación, es importante remarcar que el EGPM-CD debe ser multidisciplinar y con habilidades suficientes como para afrontar esta actividad, que puede ser considerada como una de las más importantes de la metodología presentada. La Tabla 7 muestra los principales artefactos para esta tarea.

**Tabla 7.** Artefactos para EPM.3.

|                         |         |  |
|-------------------------|---------|--|
| Productos               | Entrada | - Necesidades de información.<br>- Conceptos medibles relevantes.<br>- Modelo de datos y procesos. |
|                         | Salida  | - Lista de medidas a aplicar para cada concepto medible.   |
| Técnicas y herramientas |         | - GQM.   |
| Implicados              |         | - Responsables del EGPM-CD.  |

### EPM.4. Localizar las Entidades a Medir.

Las entidades a medir se pueden encontrar en almacenes de datos de distinta naturaleza como, por ejemplo, bases de datos relacionales o semi-estructuradas, ficheros de acceso secuencial, documentos XML u hojas de cálculo.

Para planificar la medición, es preciso localizar la ubicación de las entidades cuyo nivel de CD va a ser medido. Estas entidades tienen atributos medibles que deben ser inspeccionados a la hora

de trazar el plan de medición. Algunos ejemplos de entidades pueden ser esquemas de datos, valores de datos, dominios de datos, reglas de negocio, o interfaces de usuario [7].

Una de las principales técnicas para localizar estas entidades son las sesiones de trabajo con los responsables de los recursos de datos de la organización. En la Tabla 8, se resumen los principales artefactos para esta tarea.

**Tabla 8.** Artefactos para EPM.4.

|                         |         |  |
|-------------------------|---------|--|
| Productos               | Entrada | - Lista de entidades a medir.<br>- Modelo de datos de la organización. |
|                         | Salida  | - Lista de la ubicación de los repositorios de datos a medir.          |
| Técnicas y herramientas |         | - Inspecciones.<br>- Sesiones de trabajo.                              |
| Implicados              |         | - Componentes del EGPM-CD.<br>- Responsables de los recursos de datos. |

### **EPM.5. Determinar la Cantidad de Datos a Medir**

Dependiendo del propósito de la medición o la necesidad de no sacrificar el rendimiento del SI, puede resultar necesario delimitar el número de entidades que se deben tener en cuenta para la medición. En tal caso, se debe extraer una muestra estadísticamente representativa de todo el conjunto de entidades y luego extrapolar los resultados. Los parámetros de la muestra (ratio aceptable de datos no válidos, tamaño de la muestra, tipo de muestreo, máximo y mínimo valor para la aceptación o rechazo) pueden ser calculados de acuerdo a estándares como ISO 2859 o UNE 66020, siempre y cuando las entidades sobre las que se va a medir verifiquen las condiciones y limitaciones impuestas por estos estándares. En la Tabla 9 se pueden ver los artefactos para esta tarea.

**Tabla 9.** Artefactos para EPM.5.

|                         |         |  |
|-------------------------|---------|--|
| Productos               | Entrada | - Necesidades de información.<br>- Repositorios de datos a medir.<br>- Coste computacional necesario para medir la colección de datos completa.<br>- Relación esfuerzo-coste para la medición de la colección de datos completa. |
|                         | Salida  | - Estudio de viabilidad de la medición de la colección de datos completa.<br>- Cantidad de datos a medir para cada entidad.  |
| Técnicas y herramientas |         | - UNE-EN-ISO66020 / ISO 2859.  |
| Implicados              |         | - Componentes del EGPM-CD.   |

### **EPM.6. Cuándo y Cómo Realizar la Medición.**

La medición puede no ser una actividad puntual sino que su realización requiera un tiempo, o simplemente puede ocurrir que alguien esté interesado en estudiar la evolución temporal del nivel de CD de una entidad [31]. También es fundamental considerar la asignación temporal que se hizo durante la tarea COM.3 para evitar colisiones temporales entre los componentes del EGPM-CD que desarrollen otras labores dentro de la organización aparte de su intervención en la planificación de la medición.

Por estas razones, es necesario realizar una planificación temporal de la medición de modo que se puedan obtener los valores más significativos, y no se entorpezca el trabajo realizado por los implicados. Algunas herramientas que se pueden utilizar para la planificación temporal son los diagramas de Gantt o incluso, si se pretendiese localizar en determinados puntos del proceso de negocio, se podría utilizar alguna notación de modelado de procesos de negocio, como BPMN[30] o IPMAP, notación específica de calidad de datos desarrollada por Shankaranarayan et al. [32].



Una idea muy interesante sería automatizar los procedimientos de medición de cara a ahorrar recursos necesarios en la ejecución del plan de medición y obtener resultados más fiables.

Para poder interpretar adecuadamente los resultados, es importante describir apropiadamente la forma en que los resultados de la medición serán comunicados como parte del Plan de Medición [20, 35].

En la Tabla 10 se recogen los principales artefactos para esta tarea.

**Tabla 10.** Artefactos para EPM.6.

|                         |         |   |
|-------------------------|---------|---|
| Productos               | Entrada | <ul style="list-style-type: none"> <li>- Lista de medidas a aplicar.</li> <li>- Lista de entidades a medir.</li> <li>- Cantidad de datos a medir para cada entidad.</li> <li>- Planificación temporal de la actividad de la organización.</li> <li>- Ciclo de vida de los datos.</li> </ul> |
|                         | Salida  | <ul style="list-style-type: none"> <li>- Plan de medición.</li> </ul>   |
| Técnicas y herramientas |         | <ul style="list-style-type: none"> <li>- Diagramas de Gantt.</li> <li>- BPMN, IPMAP, Diagramas de actividad UML.</li> <li>- Sesiones de trabajo</li> </ul>  |
| Implicados              |         | <ul style="list-style-type: none"> <li>- Componentes del EGPM-CD.</li> </ul>  |

## 4 Conclusiones y Trabajo Futuro

La medición es una actividad clave en cualquier iniciativa de gestión de la calidad. Dado que las organizaciones han empezado a darse cuenta de que algunos de los problemas que sufren son debidos a niveles inadecuados en la calidad de los datos, han comenzado a dedicar esfuerzos y recursos orientados a gestionar la calidad de los datos usados en sus procesos de negocio.

En este sentido, los investigadores en el campo de la CD están comprometidos con el desarrollo de artefactos que ayuden a las organizaciones a medir su CD. Como el campo de la CD es todavía muy joven, necesita apoyarse en otros más consolidados, como el de la Ingeniería del Software, de modo que se puedan basar los esfuerzos de investigación en fundamentos sólidos, como los estándares existentes para la medición y la calidad del software. Hemos elegido uno de estos estándares, ISO/IEC 15939, como la base para la metodología MEPLAMECAL, presentada en este trabajo. Nuestra propuesta pretende llenar un vacío existente y además complementar a los trabajos de evaluación de CD ya existentes en la bibliografía.

Entendemos que la diferencia fundamental entre medición y evaluación consiste en el enfoque específico de ésta última de determinar la validez y utilidad de los datos dentro de un contexto; mientras que la medición está sólo encaminada a la obtención de valores que se utilizarán en la evaluación, sin dar soporte a juicio alguno. A pesar de que pueda parecer que esta simplificación resta relevancia a nuestra propuesta, no existe en la literatura ninguna iniciativa similar con la suficiente generalidad, y que además permita tener en cuenta las características especiales de los datos así como los aspectos inherentes a la medición de su calidad dentro del contexto organizacional.

La principal contribución de este artículo no es la metodología en sí, sino la ventaja de poder utilizarla como guía eficiente que tiene en cuenta los aspectos anteriormente citados en la planificación de la medición de la CD de las entidades organizacionales que contienen los datos que intervienen en los procesos de negocio.

MEPLAMECAL consiste en dos actividades con sus correspondientes tareas. Para facilitar el uso de la metodología, se han identificado los artefactos para cada una de estas tareas. Dado que la metodología ha sido adaptada a partir del estándar ISO/IEC 15939, muy conocido en el campo de la medición de software, cualquier profesional familiarizado podría aplicar fácilmente la metodología en su propio contexto e ir introduciendo en su catálogo de buenas prácticas los conceptos de calidad de datos.

Finalmente, es importante resaltar que para conseguir que el proceso de medición sea lo más efectivo posible es necesario que no se vea como una actividad puntual o un proceso aislado, sino como un proceso más que se integra dentro de la actividad diaria de la organización.

Somos conscientes de la importancia de obtener medidas de CD repetibles. En consecuencia, nuestra línea de trabajo actual se centra en dos aspectos: por un lado, validar la metodología mediante su aplicación a casos prácticos, con el objetivo de identificar las necesidades de información que podrían aceptar medidas con procedimientos de medición automatizables; y por otro lado, estamos desarrollando un conjunto de herramientas para la automatización del propio proceso de elaboración de Planes de Medición de CD mediante el uso de Tecnologías Semánticas.

## Agradecimientos

Esta investigación es parte de los proyectos ESFINGE (TIN2006-15175-C05-05), CALIPSO (TIN 2005-24055-E), ambos apoyados por el Ministerio de Educación y Ciencia, y HERMES (TSI-020100-2008-155) apoyado por el Ministerio de Industria, Turismo y Comercio.

## Referencias

1. Al-Hakim, L., *Procedure for Mapping Information Flow: A case of Surgery Management Process*, in *Information Quality Management: Theory and Applications*, L. Al-Hakim, Editor. 2007, Idea Group Publishing: Hershey, PA, USA. p. 168-188.
2. Batini, C. and M. Scannapieco, *Data Quality: Concepts, Methodologies and Techniques*. Data-Centric Systems and Applications. 2006, Berlin: Springer-Verlag Berlin Heidelberg.
3. Bobrowski, M., M. Marrá, and D. Yankelevich. *A Software Engineering View of Data Quality*. in *Second International Software Quality in Europe*. 1998. Brussels, Belgium.
4. Burgess, M.S.E., W.A. Gray, and N.J. Fiddian. *Quality Measures and the Information Consumer*. in *Ninth International Conference on Information Quality (ICIQ'04)*. 2004. MIT, Cambridge, MA, USA.
5. Burzynski, T. *Establishing the Environment for Implementation of a Data Quality Management Culture in the Military Health System*. in *Third International Conference on Information Quality (ICIQ'98)*. 1998. MIT, Cambridge, MA, USA.
6. Caballero, I., Ó. Gómez, and M. Piattini. *Getting Better Information Quality by Assessing and Improving Information Quality Management*. in *Ninth International Conference on Information Quality (ICIQ'04)*. 2004. MIT, Cambridge, MA, USA.
7. Caballero, I., et al. *A Data Quality Measurement Information Model based on ISO/IEC 15939*. in *12th International Conference on Information Quality*. 2007. MIT, Cambridge, MA.
8. Caballero, I., et al. *DQRDFS: Towards a Semantic Web Enhanced with Data Quality*. in *Web Information Systems and Technologies*. 2008. Funchal, Madeira, Portugal.
9. Caro, A., et al., *A proposal for a set of attributes relevant for Web Portal Data Quality*. *Software Quality Journal*, 2008.
10. DeAmicis, F., D. Barone, and C. Batini. *An Analytical Framework to analyze Dependencies among data Quality Dimensions*. in *ICIQ'06*. 2006. MIT, Cambridge, MA, USA.
11. Deming, W.E., *Out of Crisis*. 1986, Cambridge: MA: MIT Center for Advanced Engineering Study.
12. English, L., *Improving Data Warehouse and Business Information Quality: Methods for reducing costs and increasing Profits*. 1999, New York, NY, USA: Willey & Sons.
13. Eppler, M. and M. Helfert. *A Classification and Analysis of Data Quality Costs*. in *International Conference on Information Quality*. 2004. MIT, Cambridge, MA, USA.
14. Eppler, M. and P. Muenzenmayer. *Measuring Information Quality in the Web Context: A Survey of State-of-the-Art Instruments and an Application Methodology*. in *Proceeding of the Seventh International Conference on Information Quality*. 2002.
15. Franch, X. and J.P. Carvallo, *Using Quality Models in Software Package Selection*. *IEEE Software*., 2003. 20(1): p. 34-41.
16. Ge, M. and M. Helfert. *A Review of Information Quality Research*. in *Interantional Conference on Information Quality*. 2007. MIT, Cambridge, MA, USA.

17. Gebauer, M., P. Caspers, and N. Weigel. *Reproducible Measurement of Data Quality Field*. in *Tenth International Conference on Information Quality (ICIQ'05)*. 2005. MIT, Cambridge, MA, USA.
  18. Gendron, M. and M.J. D'Onofrio. *Formulation of a Decision Support Model Using Quality Attributes*. in *Seventh International Conference on Information Quality (ICIQ'02)*. 2002. MIT, Cambridge, MA, USA.
  19. Gustavsson, M. *Information Quality Measurement*. in *International Conference on Information Quality*. 2006. MIT, Cambridge, MA, USA.
  20. Humphrey, W.S., *Managing the Software Process*. The SEI Series in Software Engineering. 1989, Pittsburgh, PA Addison-Wesley.
  21. ISO-25012, *ISO/IEC 25012: Software Engineering - Software Quality Requirements and Evaluation (SQuaRE) - Data Quality Model (Draft)*. 2006.
  22. ISO/IEC, *ISO/IEC 9126. Software Engineering-Product Quality. Parts 1 to 4*. 2001: International Organization for Standardization/International Electrotechnical Commission.
  23. ISO/IEC, *ISO/IEC 15939. Information Technology - Software Measurement Process*. 2000.
  24. Lee, Y.W., et al., *Journey to Data Quality*. 2006, Cambridge, MA, USA: Massachusetts Institute of Technology.
  25. Lee, Y.W., et al., *AIMQ: A Methodology for Information Quality Assessment*. Information and Management, 2002. **40**(2): p. 133-146.
  26. Levis, M., M. Helfert, and M. Brady. *Information Quality Management: Review of an Evolving Research Area*. in *ICIQ'07*. 2007. MIT, Cambridge, MA, USA.
  27. Loshin, D., *Enterprises Knowledge Management: The Data Quality Approach*. 2001, San Francisco, CA, USA: Morgan Kaufman.
  28. Naumann, F. and C. Rolker. *Assessment Methods for Information Quality Criteria*. in *Fifth International Conference on Information Quality (ICIQ'2000)*. 2000. MIT, Cambridge, MA, USA.
  29. Oliveira, P., et al. *A Taxonomy of Data Quality Problems*. in *Second International Workshop on Data and Information Quality (in conjunction with CAISE'05)*. 2005. Porto, Portugal.
  30. OMG, *Business Process Model and Notation 2*. 2008, Object Management Group.
  31. Redman, T., *Data Quality: The field guide*. 2000, Boston: Digital Press.
  32. Shankaranarayan, G., R.Y. Wang, and M. Ziad. *IP-MAP: Representing the Manufacture of an Information Product*. in *Fifth International Conference on Information Quality (ICIQ'2000)*. 2000. MIT, Cambridge, MA, USA.
  33. Strong, D., Y. Lee, and R. Wang, *Ten Potholes in the Road to Information Quality*. IEEE Computer, 1997: p. 38-46.
  34. Strong, D.M., Y.W. Lee, and R.Y. Wang, *Data Quality in Context*. Communications of the ACM, 1997. **40**(5): p. 103-110.
  35. Ukko, J., J. Karhu, and H. Rantanen, *How to communicate measurement information successfully in small and medium-sized enterprises: a regression model*. International Journal of Information Quality, 2007. **1**(1): p. 41-59.
  36. Wang, R., et al., eds. *Information Quality*. Advances in Management Information Systems, ed. V. Zwass. 2005, M.E. Sharpe: Saddle River, NJ.
  37. Wang, R.Y., *A Product Perspective on Total Data Quality Management*. Communications of the ACM, 1998. **41**(2): p. 58-65.
  38. Wang, R.Y., M. Reddy, and H. Kon, *Towards quality data: An attribute-based approach*. Journal of Decision Support Systems, 1995. **13**(3-4): p. 349-372.
-



Universidad de Oviedo

400  
cuarto centenario



Ayuntamiento de Gijón



GOBIERNO DE ESPAÑA

MINISTERIO DE CIENCIA E INNOVACION



GOBIERNO DEL PRINCIPADO DE ASTURIAS



INTERSYSTEMS

cajAstur



Sistedes  
Sociedad de Ingeniería del Software y  
Tecnologías de Desarrollo de Software

