# ADVANCES ON INFORMATION PROCESSING AND MANAGEMENT

#### **CONFERENCE ORGANIZERS INSTITUTES**

The International Conference on Integrated Information is supported by the following Institutes:

Emerald Group Publishing Limited
Technological educational Institute of Athens, Greece
University of Peloponnese, Greece
National And Kapodistrian University of Athens, Greece
Mednet Hellas, The Greek Medical Network
2<sup>nd</sup> AMICUS Workshop

# INTEGRATED INFORMATION

International Conference on Integrated Information

Kos, Greece September, 29 – October, 3 2011

#### **EDITORS**

Georgios A. Giannakopoulos *Technological Educational Institute of Athens, Greece* 

Damianos P. Sakas *University of Peloponnese, Greece* 

## All papers have been peer-reviewed



Piraeus, Greece, 2011

#### **Editors**

Georgios A. Giannakopoulos

Technological Educational Institute of Athens Faculty of Management and Economics Department of Library Science and Information Systems Address: Aghiou Spyridonos Street, 12210, Egaleo

E-mail: gian@teiath.gr

Damianos P. Sakas

University of Peloponnese Faculty of Science and Technology Department of Computer Science and Technology Address: End of Karaiskaki St., 22100, Tripolis, Greece

E-mail: D.Sakas@uop.gr

The copyrights will be owned by the authors under the Creative Commons Attribution-Non Commercial license (<a href="http://creativecommons.org/licenses/by-nc/3.0/">http://creativecommons.org/licenses/by-nc/3.0/</a>), which permits unrestricted use, distribution, and reproduction in any non commercial medium, provided the original work is properly cited.

ISSN:

Printed in the Greece, EU

# **CONTENTS**

PREFACE: Proceedings of the International Conference on Integrated Information (IC-INFO 2011)	1
Georgios A. Giannakopoulos, Damianos P. Sakas	
Conference Details	3
Keynote Speaker	5
SYMPOSIUM ON INFORMATION AND KNOWLEDGE MANAGEMENT Prof. Christos Skourlas	6
Towards the Preservation and Availability of Historical Books and Manuscripts: A Case Study  Eleni Galiotou	8
An Extensive Experimental Study on the Cluster-based Reference set Reduction for Speeding-up the k-nn Classifier  Stefanos Ougiaroglou, Georgios Evangelidis and Dimitris A. Dervos	12
Exploiting the Search Culture Modulated by the Documentation Retrieval Applications Nikitas N. Karanikolas and Christos Skourlas	16
Information and Knowledge Organization: The Case of the TEI of Athens Anastasios Tsolakidis, Manolis Chalaris and Ioannis Chalaris	22
Providing Access to Students with Disabilities and Learning Difficulties in Higher Education through a Secure Wireless framework  Catherine Marinagi and Christos Skourlas	26
Improving Query Efficiency in High Dimensional Point Indexes Evangelos Outsios and Georgios Evangelidis	30
Text Segmentation Using Named Entity Recognition and co-Reference Resolution in Greek Texts	34
Pavlina Fragkou	
KINISIS, a Graphical XQuery Language Euclid Keramopoulos, Achilleas Pliakas, Konstantinos Tsekos and Ignatios Deligiannis	42
Dimensionality Curse, Concentration Phenomenon and the KDB-tree Nikolaos Kouiroukidis and Georgios Evangelidi	46

Applying Balanced Scorecard Strategic Management in Higher Education Manolis Chalaris, Anastasios Tsolakidis and Ioannis Chalaris	50
A Web Portal Model for NGOs' Knowledge Management Zuhal Tanrikulu	54
The Digital Archives System and Application Optimized for the Tradition Knowledge Archives	58
Jeon Hong. Chan, In Deok. Hwang, Jae Hak. Park, Hyeok. Sim, U won. Gwon and Soon Cheol. Park	
A Semi-automatic Emerging Technology Trend Classifier Using SCOPUS and PATSTAT	62
Seonho Kim, Woondong Yeo, Byong-Youl Coh, Waqas Rasheed, Jaewoo Kang	
Presenting a Framework for Knowledge Management within a Web Enabled Living Lab Lizette de Jager and Albertus AK Buitendag and Potjie (JS) van der Walt	66
4TH SYMPOSIUM ON BUSINESS AND MANAGEMENT AND DYNAMIC SIMULATION MODELS SUPPORTING MANAGEMENT STRATEGIES Dr. Damianos Sakas	71
New Political Communication Practices: No Budget Events Management. The New Challenge  Evangelia N. Markaki, Damianios P. Sakas and Theodore Chadjipantelis	73
Free Software – Open Source Software. A Powerful Tool for Developing Creativity in the Hands of the Student  Nacionavlas K. Dimitrias Demicros P. Sakas Konstantinos Massalas	78
Nasiopoulos K. Dimitrios, Damianos P. Sakas, Konstantinos Masselos	
Open Source Web Applications. How it Spread Through the Internet and their Contribution to Education.	82
Nasiopoulos K. Dimitrios, Damianos P. Sakas, Konstantinos Masselos	
Culture in Modern Times in the Frame of Luhmann's System Theory Anastasia J. Chournazidis	85
Managing Scientific Journals: A Cultural Viewpoint Marina C. Terzi, Damianos P. Sakas, and Ioannis Seimenis	87
A Conceptual Framework for Analyzing Knowledge-based Entrepreneurship Nikos S. Kanellos	92

SESSION ON INFORMATION HISTORY: PERSPECTIVES, METHODS AND CURRENT TOPICS	96
Prof. Laszlo Karvalics	
Emerging Research Fields in Information History Laszlo Z. Karvalics	98
Information Management through Elementary Data Clusters: New Observations on Pridianum-Type Roman Statistical Documents  Gergő Gellérfi	102
Information and Secrecy on the Silk Road. Methods of Encryption of Legal Documents in Inner Asia (3th-4th century)  Szabolcs Felföldi	106
The Role of Information and Disinformation in the Establishment of the Mongolian Empire: A Re-examination of the 13th century Mongolian History from the Viewpoint of Information History	110
Márton Gergő Vér	
Early Warning Systems and the Hospitallers in the Eastern Mediterranean Zsolt Hunyadi	114
Information Management as Establishment Dutch Navigational Knowledge on Japan, 1608-1641	118
Gabor Szommer	
Files Everywhere - Register and Training of Men for Military and Civil Purpose in Prussia in the early 18th century  Marton Holczer	123
SYMPOSIUM ON INTEGRATED INFORMATION: THEORY, POLICIES, TOOLS Prof. Georgios Giannakopoulos	126
Approaching Information as an Integrated Field: Educating Information Professionals Georgios Giannakopoulos, Daphne Kyriaki Manesi and Sryridon Zervos	128
Special Libraries as Knowledge Management Centers Eva Semertzaki	132
Digital Libraries' Developers and their Suitability: A Case Study Maria Monopoli	136

A Preliminary Study for the Creation of a Greek Citation index in the Humanities and the Social Sciences (GCI – H&SS)	140
Daphne Kyriaki-Manessi and Evi Sachini	
Archiving as an Information Science. Evidence from a Survey Carried out on a Sample of Greek Students	144
Georgios Giannakopoulos and Ioannis Koumantakis	
Transition Process of E-records Management and Archiving System in Universities: Ankara University	147
S. Özlem Bayram and Fahrettin Ozdemirci	
Government Information: Access and Greece's Efforts for Access Aikaterini Yiannoukakou	150
School Archives and their Potentials in Teaching: Aspects of Greek Reality Sonia Geladaki and Panagiota Papadimitriou	156
Research on School Libraries in Greece and Suggestions on its Further Development Georgios D. Bikos	160
Building Digital Collections for Archeological Sites: Metadata Requirements and CIDOC CRM Extension	164
Georgios S. Gkrous and Mara Nikolaidou	
Museological Claims to Autonomous Knowledge: Rethinking the Conceptual Mode of Display and its Claims to Knowledge	169
Assimina Kaniari and Georgios Giannakopoulos	
Use of Library Loan Records for Book Recommendation Keita Tsuji, Erika Kuroo, Sho Sato, Ui Ikeuchi, Atsushi Ikeuchi, Fuyuki Yoshikane and Hiroshi Itsumura	172
Developing a National Database on Librarianship and Information Science. The Case of E-VIVA, the Hellenic Fulltext Database Filippos Ch. Tsimpoglou, Vasiliki V. Koukounidou and Eleni K. Sakka	176
Integrated Access to Cultural Heritage Information Pieces in Iran Astan-Quds Razavi's Organization of Libraries, Museums and Documents Center: A Theory of Unionization Disparate Information Assets over Imam Reza's Zarih  Ms. Mitra Zarei and Ms. Maliheh Farrokhnia	181
Attitudes of University Librarians and Information Scientists towards the Draft Code of	185

## Library Ethics to Present a Model for Final Library Ethical Codes

Mahsoomeh Latifi, Fatemeh Zandian and Hasan Siamian

SESSION ON OPEN ACCESS REPOSITORIES: SELF-ARCHIVING, METADATA, CONTENT POLICIES, USAGE	188
Dr. Alexandros Koulouris	
Geographical Collections in Greek Academic Libraries: Current Situation and Perspectives	189
Ifigenia Vardakosta and Sarantos Kapidakis	
Information Seeking Behavior: Factors that Affect the Behavior of Greek Astronomers Hara Brindesi and Sarantos Kapidakis	194
Aggregating Metadata for Europeana: The Greek Paradigm Alexandros Koulouris, Vangelis Banos and Emmanouel Garoufallou	198
Integrating a Repository with Research Output and Publications: The Case of the National Technical University of Athens	202
Dionysis Kokkinos	
Implementation of Workflows as Finite State Machines in a National Doctoral Dissertations Archive	205
Nikos Houssos, Dimitris Zavaliadis, Kostas Stamatis and Panagiotis Stathopoulos	
Practices of "Local" Repositories of Legally Protected Immovable Monuments. A Global Scheme for 'Designation – Significance' Information  Michail Agathos and Sarantos Kapidakis	209
Sur sa a man a rea of	
Integration of Metadata in BWMETA-2.0.0 Format  Katarzyna Zamlynska, Jakub Jurkiewicz and Lukasz Bolikowski	213
SESSION ON EVIDENCE-BASED INFORMATION IN CLINICAL PRACTICE Dr. Evangelia Lappa	216
Applicability of Data Mining Algorithms on Clinical Datasets Wilfred, Bonney	218
Changing Roles of Health Librarians with Open Access Repositories Christine Urquhar and Assimina Vlachaki	221
From Medical Records to Health Knowledge Management Systems: The Coding to Health Sector	225
Evangelia C. Lappa and Georgios A. Giannakopoulos	

The Survey of Skill, Attitude and Use of Computer and Internet among Faculty Members Hasan Siamian, Azita Bala Ghafari, Kobra Aligolbandi, Mohammad Vahedi and Gholam Ali Golafshani Jooybari	229
Trends in Scholarly Communication among Biomedical Scientists in Greece Assimina Vlachaki and Christine Urquhart	232
SESSION ON ELECTRONIC PUBLISHING: A DEVELOPING LANDSCAPE Dr. Dimitris Kouis	236
E-Journal and Open Access Journal Publishing in the Humanities: Preliminary Results from a Survey among Byzantine Studies Scholars Victoria Tsoukala and Evi Sachini	238
Preliminary Results on a Printed VS Electronic Text Books Assessment Through Questionnaire  Dimitrios A. Kouis and Kanella Pouli	242
An Interpretation of Aristotelian Logic According to George Boole Markos N. Dendrinos	246
SESSION ON INFORMATION CONTENT PRESERVATION AS OUTCOME OF CONSERVATION OF CULTURAL HERITAGE: ETHICS, METHODOLOGY AND TOOLS  Prof. George Panagiaris and Dr. Spiros Zervos	251
Intrinsic Data Obfuscation as the Result of Book and Paper Conservation Interventions Spiros Zervos, Alexandros Koulouris and Georgios Giannakopoulos	254
Mass Deacidification: Preserving More than Written Information Michael Ramin, Evelyn Eisenhauer and Markus Reist	258
Information Literacy of Library Users: A Case Study of Mazandaran Public Library Users, Iran Hussein Mahdizadeh and Hasan Siamian	260
The Narratives of Paper in The Archives of the New Independent Greek State (Mid 19th c.)  Ourania Kanakari and Maria Giannikou	264
From Macro to Micro and from Micro to Nano: The Evolution of the Information Content Preservation of Biological Wet Specimen Collections Nikolaos Maniatis and Georgios Panagiaris	268

Digital images: A valuable scholar's tool or misleading material?  Patricia Engel	272
Attitudes of University Librarians and Information Scientists Towards the Draft Code of Library Ethics to Present a Model for Final Library Ethical Codes  Mahsoomeh Latifi, Fatemeh Zandianand and Hasan Siamian	277
Investigation of the Degradation Mechanisms of Organic Materials: From Accelerated Ageing to Chemometric Studies	280
Ekaterini Malea, Effie Papageorgiou and Georgios Panagiaris	
SESSION ON DIVERGENCE AND CONVERGENCE: INFORMATION WORK IN DIGITAL CULTURAL MEMORY INSTITUTIONS  Dr. Susan Myburgh	285
Extending Convergence and Divergence in Cultural Memory Institutions: The Old Slave Lodge in the New South Africa  Archie L Dick	287
The Transfer of Knowledge from Large Organizations to Small: Experiences from a Research Project on Digitization in Wales  Clare Wood-Fisher, Richard Gough, Sarah Higgins, Menna Morgan, Amy Staniforth and Lucy Tedd	289
The Usage of Reference Management Software (Rms) in an Academic Environment : A Survey at Tallinn University  Enrico Francese	293
Varialog : How to Locate Words in a French Renaissance Virtual Library Marie-Hélène Lay	297
The Urge to Merge: A Theoretical Approach Susan Myburgh	301
SYMPOSIUM ON ADVANCES INFORMATION FOR STRATEGIC MANAGEMENT Professor Nikolaos Konstantopoulos	304
Empowerment in the Tax Office of Greece Antonios E. Giokas and Nikolaos P. Antonakas	306
Building Absorptive Capacity Through Internal Corporate Venturing Ioannis M. Sotiriou and Alexandros I. Alexandrakis	310

The Monitoring Information System (M.I.S.) - An information and Management System for Projects Co-financed Under the National Strategic Reference Framework (NSRF) and the Community support framework (CSF)	
Catherina G. Siampou, Eleni G. Fassou and Athanassios P. Panagiotopoulos	
Corruption in Tax Administration: The Entrepreneurs View Point	318
Nikolaos P. Antonakas, Antonios E. Giokas and Nikolaos Konstantopoulos	
Conflicts between the IT Manager and the Software House after the Strategic Choice of Outsourcing of the Information Processes in Maritime Companies.	322
Anthi Z. Vaxevanou, Nikolaos Konstantopoulos, Damianos P. Sakas	
Contemporary Forms of Ordering Between the Supply Department and Ship Chandler Companies in the Shipping Industry	325
Anthi Z. Vaxevanou, Nikolaos Konstantopoulos, Damianos P. Sakas	
Strategies Implemented and Sources Used for the Acquisition of Information on Foreign Markets	329
Myropi Garri, Nikolaos Konstantopoulos and Michail G. Bekiaris	
The Effect of High Performance Working Systems on Informative Technology in Enterprises after Organisation Changes such as Mergers & Acquisitions	333
Nikolaos Konstantopoulos and Yiannis Triantafyllopoulos	
Personnel's Absorptive Capacity as a Guiding Concept for Effective Performance in Informative Technology	337
Nikolaos Konstantopoulos and Yiannis Triantafyllopoulos	
SESSION ON CONTEMPORARY ISSUES IN MANAGEMENT: ORGANISATIONAL BEHAVIOUR, INFORMATION TECHNOLOG, EDUCATION & HOSPITAL LEADERSHIP	341
Dr. Panagiotis Trivellas	
Investigating the Importance of Sustainable Development for Hotel SMES Panagiotis Reklitis and Anestis Fotiadis	343
Strategic Alignment of ERP, CRM and E-business: A Value Creation Catherine C. Marinagi and Christos K. Akrivos	347
The Impact of Occupational Stress on Performance in Health Care Panagiotis Trivellas Panagiotis Reklitis and Charalambos Platis	351

The Impact of Emotional Intelligence on Job Outcomes and Turnover Intention in Health Care	
Panagiotis Trivellas Vassilis Gerogiannis and Sofia Svarna	
SYMPOSIUM ON BUSINESS MANAGEMENT AND COMMUNICATION STRATEGIES SUPPORTING DECISION MAKING PROCESS IN TOURISM SECTOR	360
Dr. Panagiota Dionysopoulou	
The Human Factor as a Mediator to the Total Quality in the Tourism Companies. The impact of Employees' Motivation to Quality Improvements	362
Christos K. Akrivos and Panagiotis Reklitis	
Tourist Destination Marketing and Management Using Advanced ICTS Technologies Anastasia Argyropoulou, Panagiota Dionyssopoulou, Georgios Miaoulis	365
G.N.T.O. (Greek National Tourism Organization) Communication Strategy in Advertising Campaigns 1991-2006	370
George Stafylakis and Panagiota Dionyssopoulou	
GENERAL PAPERS	375
The role of Environmental Education within the Framework of the Environmental	376
Policy of a Regional Municipality	
Vassiliki Delitheou and Dimitra Thanasia	
Issues of Social Cohesion: A case study from the Greek Urban Scenery Evgenia Tousi	380
Merging Activity and Employee Performance: The Greek Banking System Panagiotis Liargovas and Spyridon Repousis	384
Sustainable Development and Corporate Social Responsibility in Higher Education: Some	387
Evidence from Greece	
Anastasios Sepetis and Fotios Rizos	
Exploring the Effects of Organizational Culture on Collaborative vs. Competitive	
Knowledge Sharing Behaviors	395
Hanan Abdulla Mohammed Al Mehairi and Norhayati Zakaria	

# Preface: Proceedings of the International Conference on Integrated Information (IC-ININFO 2011)

#### GEORGIOS A. GIANNAKOPOULOS

Department of Library Science and Information Systems, Technological Educational Institute of Athens, Aghiou Spyridonos, Egaleo, 12210, Greece

#### DAMIANOS P. SAKAS

Department of Computer and Technology Science, University of Peloponnese, Praxitelous 89-91, Piraeus, 18532, Greece

#### Aims and Scope of the Conference

The International Conference on Integrated Information 2011 took place in Kos Island, Greece, between September, 29 and October, 3, 2011. IC-ININFO is an international interdisciplinary conference covering research and development in the field of information management and integration.

The conference aims at creating a forum for further discussion for an Integrated Information Field incorporating a series of issues and/or related organizations that manage information in their everyday operations. Therefore, the call for papers is addressed to scholars and/or professionals of the fields of Library and Archives Science (including digital libraries and electronic archives), Museum and Gallery Studies, Information Science, Documentation, Information Management, Records Management, Knowledge Management, Data management and Copyright experts the latter with an emphasis on Electronic Publications. Furthermore, papers focusing on issues of Cultural Heritage Management and Conservation Management are also be welcomed along with papers regarding the Management of Nonprofit Organizations such as libraries, archives and museums.

One of the primary objectives of the IC-ININFO will be the investigation of information-based managerial change in organizations. Driven by the fast-paced advances in the Information field, this change is characterized in terms of its impact on organizations that manage information in their everyday operations.

Grouping emerging technologies in the Information field together in a close examination of practices, problems and trends, IC-ININFO and its emphases on integration and management will present the state of the art in the field. Addressed jointly to the academic and practitioner, it will provide a forum for a number of perspectives based on either theoretical analyses or empirical case studies that will foster dialogue and exchange of ideas.

### **Topics of general Interest**

Library Science, Archives Science, Museum and Gallery Studies, Information Science, Documentation, Digital Libraries, Electronic Archives, Information Management, Records / Document Management, Knowledge Management, Data Management, Copyright, Electronic Publications, Cultural Heritage Management, Conservation Management, Management of Nonprofit Organizations, History of Information, History of Collections, Health Information

## Symposia

The Conference offered a number of sessions under its patronage, providing a concise overview of the most current issues and hands-on experience in information-related fields.

- Symposium on Integrated information: Theory, Policies, Tools
- 4th Symposium on Business and Management and Dynamic Simulation Models supporting management strategies

- Session on Open Access Rrepositories: Self-archiving, Metadata, Content policies, Usage
- Session on Evidence-Based Information in Clinical Practice
- Session on Business Management and Communication Strategies supporting Decision Making Process in Tourism Sector
- Session on Electronic Publishing: A Developing Landscape
- Session on Information and Knowledge Management
- Session on Information Content Preservation as Outcome of Conservation of Cultural Heritage: Ethics, Methodology and Tools
- Session on Advances Information for Strategic Management
- Session on Information History: Perspectives, Methods and Current Topics
- Session on Divergence and Convergence: Information Work in Digital Cultural Memory Institutions
- Session on Contemporary issues in Management: Organisational Behaviour, Information Technology, Education & Hospital leadership.

The wide range of aspects that the sessions covered, highlighted future trends in the Information Science.

#### **Paper Peer Review**

More than 300 papers had been submitted for consideration in IC-ININFO 2011. From them, 91 were selected for presentation, after peer review in a double blind review process. The accepted papers were presented at IC-ININFO 2011.

#### **Thanks**

We would like to thank all members that participated in any way in the IC-ININFO 2011 Conference and especially:

- The famous publishing house Emerald for its communication sponsorship.
- The co-organizing Universities and Institutes for their support and development of a high-quality Conference scientific level and profile.
- The members of the Scientific Committee that honored the Conference with their presence and provided a significant contribution to the review of papers as well as for their indications for the improvement of the Conference.
- All members of the Organizing Committee for their help, support and spirit participation before, during and after the Conference.
- The Session Organizers for their willing to organize sessions of high importance and for their editorial work, contributing in the development of valued services to the Conference.
- PhDc Marina Terzi for her excellent editorial work, contributing in the production of the Conference proceedings.

#### CONFERENCE DETAILS

#### Chairs

Georgios A. Giannakopoulos, Technological Educational Institute of Athens, Greece Damianos P. Sakas, University of Peloponnese, Greece

#### **Co-Chairs**

Daphne Kyriaki – Manesi, Technological Educational Institute of Athens, Greece Dimitrios Vlachos, University of Peloponnese, Greece

#### **Scientific Committee**

Amanda Spink, Queensland University of Technology

Andreas Bagias, European Court

Andreas Rauber, Vienna University of Technology

Astrid van Wesenbeeck, SPARC Europe

Christine Urquhart, Aberystwyth University

Christos Schizas, University of Cyprus

Christos Skourlas, Technological Educational Institute of Athens

Claire Farago, University of Colorado at Boulder

Claus-Peter Klas, FernUniversität in Hagen

Costas Vassilakis, University of Peloponnese,

Dimitris Dervos, Technological Educational Institute of Thessaloniki

Eelco Ferwerda, OAPEN

Elena Garcia Barriocanal, University of Alcalá

Emmanouel Garoufallou, Technological Educational Institute of Thessaloniki

Filippos Tsimpoglou, University of Cyprus

Fillia Makedon, University of Texas at Arlington

George Korres, University of Newcastle

Georgios Evangelidis, University of Macedonia

Georgios Panagiaris, Technological Educational Institute of Athens

Johan Oomen, Netherlands Institute for Sound and Vision

José Aldana, University of Malaga

Konstantinos Masselos, University of Peloponnese

Luciana Duranti, The University of British Columbia

Markos N. Dendrinos, Technological Institute of Athens

Milena Dobreva, University of Strathclyde

Prodromos Tsiavos, London School of Economics and Political Science

Sándor Darányi, University of Borås

Sarantos Kapidakis, Ionian University

Sirje Virkus, Tallinn University

Spiros Zervos, Technological Educational Institute of Athens

Susan Myburgh, University of South Australia

Theodoros Pitsios, University of Athens, Faculty of Medicine

#### **Organizing Committee**

Alexandros Koulouris (Chair), Technological Educational Institute of Athens

Christos Christopoulos, SCEV Scientific Events Ltd

Marina Terzi, University of the Aegean, Greece

Evangelia Markaki, Aristotle University of Thessaloniki

Assimina Kaniari, Athens School of Fine Arts Evangelia Lappa, General Hospital Attikis K.A.T. Dimitris Kouis, Greek Ministry of Education, Lifelong Learning and Religious Affairs Dionysis Kokkinos, National Technical University of Athens

#### KEYNOTE SPEAKER



Professor Amanda Spink

Professor Amanda Spink has published over 340 scholarly journal articles, refereed conference papers and book chapters, and 6 books. Many of her journal articles are published in the Journal of the American Society for Information Science and Technology, Information Processing and Management, and the Journal of Documentation. She is Editor of the Emerald journal Aslib Proceedings. Amanda's research has been published at many conferences including ASIST, IEEE ITCC, CAIS, Internet Computing, ACM SIGIR, and ISIC Conferences. Her recent books include Information Behavior: An Evolutionary Instinct and Web Search: Multidisciplinary Perspectives, both published by Springer. Amanda's research focuses on theoretical and empirical studies of information behavior, including the evolutionary and developmental foundations. The National Science Foundation, the American Library Association, Andrew R. Mellon Foundation, Amazon.com, Vivisimo. Com, Infospace.com, NEC, IBM, Excite.com, AlltheWeb.com, AltaVista.com, FAST, and Lockheed Martin have sponsored her research. In 2008 Professor Spink had the second highest H-index citation score in her field from 1998 to 2008 [Norris, M. (2008)]. Ranking Fellow Scholars and their H-Index: Preliminary Survey Results. Loughborough University, Dept of Information Science Report].

# Investigation of the Degradation Mechanisms of Organic Materials: From Accelerated Ageing to Chemometric Studies

Ekaterini Malea †, Effie Papageorgiou ‡ and Georgios Panagiaris †

† Technological Educational Institute of Athens, Department of Conservation of Antiquities & Works of Art, Ag. Spyridonos str., GR- 122 10 Aegaleo, Greece.

kmalea (at) teiath.gr gpanag (at) teiath.gr

‡ Technological Educational Institute of Athens, General Department of Mathematics, Ag. Spyridonos str., GR-122 10 Aegaleo, Greece.

Efipapag (at) teiath.gr

Abstract: The most common way followed for the determination of degradation mechanisms of organic materials is the application of physicochemical methods after accelerated ageing. The disadvantage of this practice is the inability to control the effect of the combination of degradation factors in real time and in real conditions as it happens in burial environment, in museum environment or in open-air conditions. In recent years it has been noticed a remarkable research activity towards the development of alternative decay determination methods in organic materials based on statistics (Chemometric studies). The discipline of Conservation of Antiquities and Works of Art has not developed yet its own statistic tools therefore it borrows the statistical methods used in social sciences and biosciences. One could claim that biostatistics is the most appropriate to answer the questions of the Conservators because faces similar "cases". At the same time it is necessary to establish a network of interested conservators/restorers and scientists and to build a central database containing all experimental data and software for all members of the network.

**Keywords:** Conservation, Organic materials, Chemometric studies.

#### I. INTRODUCTION

Main concern of the Conservator's work is the safeguarding of the aesthetic, historical, structural and material integrity of the objects he/she conserves. This is achieved by the diagnosis of their state of preservation, by planning and applying the methods of preventive and interventive conservation. The cultural properties' integrated information content is related not only to their historical identity but also to the artefacts' technology, their material structure and their behavior towards the environment during natural ageing. The methods and the materials used in conservation should be compatible with the research that follows in order to highlight the historical identity of the cultural evidence. (E.C.C.O., 2002)

The investigation of the degradation mechanisms of the materials constituting the Cultural and Natural Heritage, despite the progress made, has a long way still to go. The most common way followed for the determination of these mechanisms is the application of physicochemical methods after accelerated ageing. The disadvantage of this practice is the inability to control the effect of the combination of degradation factors in real time and in real conditions as it happens in burial environment, in museum environment or in open-air conditions. This problem is getting even worse in the case of organic materials that are characterized by a variety in structure and composition, and that are extensively sensitive in many environmental factors, abiotic and – mainly– biotic.

The need for understanding the decay process - especially of these organic materials with industrial use (i.e. paper, wood) - resulted to the standardization of diagnostic techniques and of the accelerated ageing protocols (ASTM, 1993; ASTM, 1995; ASTM, 1996; ASTM, 2002a; ASTM, 2002b). Although this standardization, it is no coincidence that in recent years it has been noticed a remarkable research activity towards the development of alternative decay determination methods in organic materials based on statistics.

This method is based on statistical elaboration of analytical data from groups of existent objects and aims at the correlation of the factors that –potentially– are responsible for their decay and at the determination of the critical factor that can be considered as catalyst for their preservation (Malea, 1999; Larsen *et al.*, 2002; Bechmann and Larsen, 2007; Zervos, 2007; Deselnicu, 2010; Malea, 2010).

#### II. METHODS

Statistics have been proved a useful tool for answering questions in many disciplines. In practice through an interactive process one studies the data in hand, examines it using some statistical method, decides to look at it another way by either transforming them, taking only a part of them or by even applying a different analysis method. Different methods are being used to bring out different aspects of the data, to ask and answer different questions about the "population" they came from and predict accurately future observations (make inference) by minimizing the time and effort needed. By using statistics the researcher has sound methodological tools that enable him to make accurate conclusions on the underlying process that generates the data itself (Berthold and Hand, 1998).

According to the generic definition, accepted by The International Chemometrics Society (ICS), chemometrics is "...the science of correlating, by statistic-mathematic techniques, of measurements conducted on a chemical system process, its state being in given conditions or at a given time". Although the term of chemometrics has been introduced in 1972 (Wold, 1972), the operative definition of chemometrics, stated by practitioners, was established in 1986 and presented in the editorial from the first issue of Chemometrics and Intelligent Laboratory Systems Journal, in the following sentence: "Chemometrics is that discipline in the field of chemistry which uses methods of mathematics and statistics to formulate or select experimental plans and procedures of physicalchemical analysis, in view of extracting maximum information by algorithmic processing of data obtained as a result of studies conducted at laboratory or industrial level" (Deselnicu, 2010).

The objective of chemometric techniques is that of highlighting complex and non-obvious relationships ("hidden") between a large number of variables associated with studied systems, resorting to statistic-mathematic models capable of describing states of physical-chemical systems, at a given time or in their dynamic.

Techniques of modelling chemometric data can be classified into:

- quantitative, regression techniques, that provide numeric parameters of postulated models;
- qualitative techniques, of recognising models/patterns/structures of multicriterial grouping of experimental data, providing images and sometimes numerical indexes of similarity/dissimilarity, on the basis of which classes and/or filiations of studied systems or of states of a system going through a given process can be identified (Deselnicu, 2010).

The discipline of Conservation of Antiquities and Works of Art has not developed yet its own statistic tools therefore it borrows the statistical methods used in social sciences and biosciences. One could claim that biostatistics is the most appropriate to answer the questions of the Conservators because faces similar "cases" such as:

- Great number of data with many missing values
- Elaboration of data derived by small number of samples tested, in relation to the large number of variables (data), correspondingly to medical discipline when investigating rare diseases.

Missing data in clinical trials reduces the power of a study and is a major source of potential bias when interpreting study results, especially if the amount of missing data is substantial. There are many reasons for missing data. There is no universally acceptable method for handling missing data. When data are missing, any attempt to draw conclusions from a statistical analysis

rests on untestable assumptions concerning the relationship between the missing (unobserved) data and the reasons why the data is missing (the missing data mechanism). The causes for missing data can be grouped into three different classes, depending on the relationship between the unobserved data and the missing data mechanism:

- a) Missing Completely at Random (MCAR). MCAR means that the missing data mechanism is unrelated to the values of any variables, whether missing or observed. In other words when we say that data are missing completely at random, we mean that the probability that an observation (X<sub>i</sub>) is missing is unrelated to the value of X<sub>i</sub> or to the value of any other variables. Thus data on family income would not be considered MCAR if people with low incomes were less likely to report their family income than people with higher incomes. Similarly, if Whites were more likely to omit reporting income than African Americans, we again would not have data that were MCAR because missingness would be correlated with ethnicity. However if a participant's data were missing because he was stopped for a traffic violation and missed the data collection session, his data would presumably be missing completely at random. Also, data that are missing because a researcher forgot to perform a test are likely to be MCAR. Another way to think of MCAR is to note that in that case any piece of data is just as likely to be missing as any other piece of data. Unfortunately, most missing data are not MCAR.
- b) Missing at Random (MAR). MAR means that the missing data mechanism is unrelated to the missing values but is related to either observed covariates or response variables. The data can be considered as missing at random if the data meet the requirement that missingness does not depend on the value of Xi after controlling for another variable. For example, people who are depressed might be less inclined to report their income, and thus reported income will be related to depression. Depressed people might also have a lower income in general, and thus when we have a high rate of missing data among depressed individuals, the existing mean income might be lower than it would be without missing data. However, if, within depressed patients the probability of reported income was unrelated to income level, then the data would be considered MAR, though not MCAR. Also, data that are missing because a subject is removed from a study based on a pre-defined clinical condition like an elevated blood glucose level are MAR.
- c) Non-Ignorable (NI). NI is also known as Missing Not at Random (MNAR). It means that the missing data mechanism is related to the missing values. For example, if we are studying mental health and people who have been diagnosed as depressed are less likely than others to report their mental status, the data are not missing at random. Clearly the mean mental status score for the available data will not be an unbiased estimate of the mean that we would have obtained with complete

data. The same thing happens when people with low income are less likely to report their income on a data collection form.

A key distinction is whether the mechanism is ignorable (i.e., MCAR or MAR) or non-ignorable.

There are excellent techniques for handling ignorable missing data. A nice feature of data that are MCAR is that the analysis remains unbiased. We may lose power for our design, but the estimated parameters are not biased by the absence of data.

About the MAR data, the phraseology is a bit awkward because we tend to think of randomness as not producing bias, and thus might well think that Missing at Random is not a problem. Unfortunately is a problem, although in this case we have ways of dealing with the issue so as to produce meaningful and relatively unbiased estimates. For data that are missing at random, Chen and Little (1999, 2001) discuss methods for analyzing survival data with missing covariates. Little and Yau (1996) develops a multiple imputation method for intent-to-treat analysis of repeated measures data with drop-outs. The SAS code used to generate the multiple imputes in the example in this paper can be accessed by clicking on Little and Yau intent-to-treat code [1]. Little and Raghunathan (1999) compare likelihood and summary approaches to longitudinal data with drop-outs in a simulation study. Likelihood-based methods for estimating the complier-average causal effects of treatments are considered in Little and Yau (1998), Yau and Little (2001). Ezzati-Rice et al. (1995) discusses a recent large-scale application of multiple imputation to a national survey. Little and An (2003) propose a robust likelihood-based method for multivariate data with missing values based on regressions on splines of propensity scores.

When we have data that are NI (or MNAR) we have a problem. The only way to obtain an unbiased estimate of parameters is to model missingness. In other words we would need to write a model that accounts for the missing data. That model could then be incorporated into a more complex model for estimating missing values. This is not a task anyone would take on lightly. See Dunning and Freedman (2008) for an example. Generally, the non-ignorable missing data are more challenging and require a different, more complicated approach. Depending on the data, this approach may include multiple imputation methods, survival analysis techniques, and pattern mixture models. For data that are not missing at random, Little (1993) discusses pattern-mixture models, a broad class of models that they do not require precise specification of the missingdata mechanism. Tang, Little and Raghunathan (2003) develop a pseudo-likelihood method for fitting no randomly missing data that avoids specifying the precise form of the mechanism. Little (1995) develops a model-based framework for repeated-measures data with drop-outs, and places existing literature within this framework.

Recently, in the United States, the Division of Behavioral and Social Sciences and Education under National Research Council of the National Academies have been working on a project "Handling missing data in clinical trials"[2]. The working group recently makes its report available. The report is titled: "The prevention and treatment of missing data in clinical trials" (National Research Council of the National Academies, 2010). Respectively we should link the cases of missing values met in Conservation of Cultural Properties with the missing data mechanisms. In example, missing data mainly of archaeological organic materials may due to loss of material, to insufficient or inadequate of incomplete diagnostic methods, or/and to documentation (loss of information) during the steps that take place between the excavation and the transportation to the Laboratory. We have to determine which of the above mentioned cases can be considered as MACR, MAR or MNAR.

Once the missing data has happened, there is no universal method to handle the missing data perfectly. The assumptions of MACR, MAR, and MNAR can never been fully verified.

Regarding the handling of the rare diseases, although there is no a single definition, the European Commission (Health and Consumer Protection Directorate-General) considers that a disease is rare if its prevalence is lower than 50 cases per 100,000 habitants. Hence, according to this definition, rare diseases will have very low observed cases in low populated areas. This definition could also be applied in archaeological organic materials that rarely can be found due to their vulnerability in degradation factors.

If the disease is rare, many areas may have zero counts, which will cause additional problems in the statistical analysis due to the large number of zeros that appear. Although standard methods could be used, care should be taken and other statistical methods that account for the excess of zeros may be required Lee et al. (2006).

Regarding the detection of disease clusters, many methods have been proposed so far and a comprehensive review can be found in Kulldorff (2006). The description of the geographical pattern of disease should be considered as a previous step to the detection of disease clusters. Regions of high risk will probably appear when displaying the relative risks in a map, but some other methods may be required to identify the actual location (and shape) of the clusters. Wakefield et al. (2000) describe some methods for the detection of disease clusters and show some examples of their use. Besag and Newell (1991) divide the types of methods for the detection of disease clusters in two groups: • General methods and • Focused methods. Another type of methods is that of Scan methods. Equally, relevant methods should be developed for the statistical elaboration of the data derived from organic materials. These data could be linked with "charts" of burial or museum environments.

The development of treatments for rare diseases can be challenging given the complexities and uncertainties regarding these diseases. To manage the effects of small patient populations and variable, irreversible disease requires the use of optimal statistical approaches, alternative study designs and for some diseases, the use of biomarkers. Industry, academia, the Food & Drug Administration, the National Institutes of Health, and other stakeholders must work together to develop strategies, policies and guidance to help improve the predictability and efficiency of the development process to enable more approvals for rare disease treatments. In other words, multidisciplinary research groups of experts in Conservation, in Material Science and in Statistics should be joined together in order to develop strategies, policies and methods of data elaboration of organic materials consisting our cultural heritage.

#### III. CONCLUSIONS

Statistics as applied in Chemometric studies opens new research horizons in the field of Conservation of Cultural and Natural Heritage and of Material Science, creating new perspective for the understanding of the materials' behavior –especially of the organic onesunder complex environmental factors. The results of statistics could then be used to create models of accelerated ageing for a better simulation of real condition effects. To achieve this, experts of statistics should develop statistical methods applied to conservation.

As stated by Larsen *et al.* (2002) a successful outcome will very much depend on the amount of data available for model building and the subsequent correlation analysis. It is important that many different techniques –advanced as well as simple- are applied to a large number of samples in different states of preservation. This will require extensive cooperation between both conservators/restorers and scientists, and an efficient flow of data and software between the different research groups. One way of achieving this may be, as in the case of IDAP (Bechmann and Larsen, 2007), to establish a network of interested conservators/restorers and scientists and to build a central database containing all experimental data and software for all members of the network.

The latter is enhanced by the very strong urge of E.U. for free access to results of research conducted in Europe.

#### REFERENCES

- ASTM D 2244-93: Standard test method for calculation of colour differences from instrumentally measured colour coordinates (1993).
- ASTM Annual book of ASTM Standards, Section 4
  Construction. 4.10 Wood, D 198: 74-75 (1995).
- ASTM E 313-96: Standard Practice for Calculating
  Yellowness and Whiteness Indices from
  instrumentally measured Colour Coordinates
  (1996).

- ASTM D 6819-02: Standard test method for accelerated temperature aging of printing and writing paper by dry oven exposure apparatus (2002a).
- ASTM D 6833-02: Test method for accelerated pollutant aging of printing and writing paper by pollution chamber exposure apparatus (2002b).
- Bechmann, D. J. and R. Larsen, "The IDAP website —
  Data collection and sharing of knowledge", in R.
  Larsen (ed.), School of Conservation, The Royal
  Academy of Fine Arts, Improved damage
  assessment of parchment IDAP, Assessment, data
  collection and sharing of knowledge, Research
  Report no 18, Copenhagen, European
  Commission-Directorate General for Research IEnvironment, 13-16 (2007).
- Berthold, M. and D. Hand, "Intelligent Data Analysis", Berthold, M. and Hand, D. (eds.), Springer-Verlag publications, Heidelberg (1998).
- Besag, J. and J. Newell, "The detection of clusters in rare diseases," Journal of the Royal Statistical Society Series A, 154:143–155 (1991)
- Chen, H. Y. and R.J.A. Little, "A Test of Missing Completely at Random for Generalized Estimating Equations with Missing Data". Biometrika, 86 (1): 1-13 (1999).
- Chen, H. Y. and R.J.A. Little, "Proportional Hazards Regression with Missing Covariates," Journal of the American Statistical Association, 94: 896-908 (1999).
- Chen, H.-Y. and R.J. Little, "A Conditional Profile

  Likelihood Approach for the Semiparametric

  Transformation Regression Model with Missing

  Covariates," Lifetime Data Analysis, 7: 207-224

  (2001).
- Deselnicu D. C., "Applications of statistical methods in investigating heritage leather items," Revista Leather and Footwear Journal, eds. INCDTP-Division: Leather and Footwear Research Institute, March 2010, Vol. 10, No 1, 13-30 (2010).
- Dunning, T., and D.A. Freedman, "Modeling selection effects," In Social Science Methodology. Sage, Steven Turner and William Outhwaite (eds), 225–231 (2007).
- E.C.C.O. Professional Guidelines 1, http://www.eccoeu.org/about-e.c.c.o./professional-guidelines.html (accessed 15/8/2011) (2002).
- Ezzati-Rice, T., W. Johnson, M. Khare, R. Little, D.

  Rubin, and J. Schafer, "A Simulation Study to

  Evaluate the Performance of Model-Based

  Multiple Imputations in NCHS Health

  Examination Surveys". Invited Paper. Annual

  Research Conference, U.S. Bureau of the Census,

  (1995).
- Kulldorff, M., "Tests of spatial randomness adjusted for an inhomogeneity: a general framework," Journal of the American Statistical Association, 101(475):1289–1305 (2006).

- Larsen, R., D. V. Poulsen, M. Odlyha, N. Kurt, J. Wouters, L. Puchinger, P. Brimblecombe and D. Bowden, "The Use of Complementary and Comparative Analysis in Damage Assessment of Parchments", in R. Larsen (ed.), Microanalysis of Parchment, London, Archetype, 165-179 (2002).
- Lee, A.H, K. Wang, J.A. Scott, K.K. Yau and G.J.

  McLachlan, "Multi-level zero-inflated poisson regression modelling of correlated count data with excess zeros," Statistical Methods in Medical Research, 15: 47–61 (2006).
- Little, R.J.A., "Pattern-mixture models for multivariate incomplete data," Journal of the American Statistical Association, 88: 125-134 (1993).
- Little, R.J.A., "Modeling the Drop-Out Mechanism in Longitudinal Studies," Journal of the American Statistical Association, 90: 1112-1121 (1995).
- Little, R.J.A. and L. Yau, "Intent-to-Treat Analysis in Longitudinal Studies with Drop-Outs," Biometrics, 52: 1324-1333 (1996).
- Little, R.J.A. and L. Yau, "Statistical Techniques for Analyzing Data from Prevention Trials: Treatment of No-Shows Using Rubin's Causal Model," Psychological Methods, 3 (2): 147-159 (1998).
- Malea, E. and D. E. Watkinson, "Contribution of analytical methods in the estimation of the condition of waterlogged archaeological leather".

  Advanced Study Course 1999, 6-10 July 1999, School of Conservation, Royal Academy of Fine Arts, Copenhagen, Denmark, European Commission Directorate-General XII, Environment and Climate Programme, Preprints 249-259 (1999).
- Malea, E., T. Vogiatzi and D. E. Watkinson, "Assessing the physical condition of waterlogged archaeological leather", at the 11th ICOM-CC WOAM WG Conference in Greenville, North Carolina USA, May 24 28, 2010, peer review paper Proc. in press.
- National Research Council, "The Prevention and Treatment of Missing Data in Clinical Trials", Panel on Handling Missing Data in Clinical Trials; The National Academies Press, (2010).
- Tang, G., R.J. Little and T. Raghunathan, "Analysis of Multivariate Monotone Missing Data by a Pseudo-Likelihood Method," Proceedings of the 2nd Seattle Symposium in Biostatistic, New York: Springer Verlag (2003).
- Wakefield, J.C., J.E. Kelsall and S.E. Morris "Clustering, cluster detection and spatial variation in risk", In Elliot P., Wakefield J.C., Best N.G. and Briggs D.J. (eds), Spatial Epidemiology. Methods and applications, chapter 8, Oxford University Press, Oxford, 128–152 (2000).
- Wold, S., "Spline-funktioner-ett nytt verktyg I dataanalysen," Svensk Kemisk Tidskrift, 1972, 3, 34-37 (1972).

- Yau, L. and R.J.A. Little, "Inference for the Complier-Average Causal Effect from Longitudinal Data Subject to Noncompliance and Missing Data, with Application to a Job Training Assessment for the Unemployed," Journal of the American Statistical Association, 96, 1232-1244 (2001).
- Zervos, S., "Evaluating treatments of paper using statistically valid test methods. Part II: Experimental setup and protocol," Restaurator 28(4): 256-288 (2007).

#### **WEBSITES**:

- [1] http://sitemaker.umich.edu/rlittle/files/linda.htm [Accessed 14 August 2011]
- [2] http://www8.nationalacademies.org/cp/projectview. aspx?key=49049 [Accessed 14 August 2011]