COVER SHEET

Title: *Integration of Data Mining Operations for Structural Health Monitoring for* Proceedings of the **8**[th] **International Workshop on Structural Health Monitoring 2011**

Authors (names are for example only): Erik Sonnleitner
Hilda Kosorus
Stefan Anderlik
Reinhard Stumptner
Bernhard Freudenthaler
Helga Allmer
Josef Küng

PAPER DEADLINE: **May 15, 2011**

PAPER LENGTH: **8 PAGES MAXIMUM **

SEND PAPER TO: **Professor Fu-Kuo Chang**
Organizing Chairman
Aeronautics and Astronautics Department
Stanford University
Durand Building
Stanford, California 94305, U.S.A.

Tel: 650/723-3466
Fax: 650/725-3377
E-mail: fkchang@stanford.edu

Please submit your paper in Microsoft Word® format or PDF if prepared in a program other than MSWord. We encourage you to read attached Guidelines prior to preparing your paper— this will ensure your paper is consistent with the format of the articles in the CD-ROM.

ABSTRACT

Data Integration is a well elaborated scientific area and one of the most important use cases of the Semantic Web. Techniques developed in this field aim at providing interoperability between heterogeneous data sources. Compared to typical Semantic Web use cases, data integration issues are manifold and also affect applications through their underlying schemas. Civil engineers specialized in risk and measurement analysis need a reliable Decision Support System (DSS) that integrates various required techniques. Hence, Structural Health Monitoring (SHM) applications tend to adopt typical integration concepts, but not by regarding data and their semantics independent from the application domain. Instead, such a DSS should be accessible in an integrated manner to support the usage of methods and techniques from different systems according to their intended operational purpose. This paper presents some practical examples of using Data Mining operations which enable a better understanding of the analysed data and which can be successfully integrated into a unified DSS for SHM.

INTRODUCTION

Standardization and Integration are two major topics in the EU funded FP7 project IRIS (Integrated European Industrial Risk Reduction System). Based on these two main objectives, a concept for an integrative usage of different measurement analyzing functions for Structural Health Monitoring (SHM) like Fast Fourier Transformation (FFT), Wavelet Transformation, etc. is required. These operations are partly available within systems like BRIMOS® (Bridge Monitoring System) (see [1] and [2]), but not in an integrative and guided manner. The problem of using such systems is the lack of semantic behaviors and conditions for measurement analysis: Many operations are available, but not everyone knows which function should be executed in which scope

Erik Sonnleitner, FAW – Institute for Applied Knowledge Processing, Johannes Kepler University Linz, Altenberger Straße 69, 4040 Linz, Austria.

Hilda Kosorus, FAW – Institute for Applied Knowledge Processing, Johannes Kepler University Linz, Altenberger Straße 69, 4040 Linz, Austria

Stefan Anderlik, DKE – Institut für Wirtschaftsinformatik - Data & Knowledge Engineering, Johannes Kepler University Linz, Altenberger Straße 69, 4040 Linz, Austria

Reinhard Stumptner, FAW – Institute for Applied Knowledge Processing, Johannes Kepler University Linz, Altenberger Straße 69, 4040 Linz, Austria

Bernhard Freudenthaler, SCCH - Software Competence Center Hagenberg GmbH, Softwarepark 21, 4232 Hagenberg, Austria

Helga Allmer, VCE Holding GmbH, Hadikgasse 60, 1140 Wien, Austria

Josef Küng, FAW – Institute for Applied Knowledge Processing, Johannes Kepler University Linz, Altenberger Straße 69, 4040 Linz, Austria

(e.g. clustering operations can be useful in addition to certain pre-existence data discovery methods) and how (parameters, semantics, notification, etc.). Therefore the integration approach of heterogeneous Decision Support Systems (DSS) for SHM in [3] should provide a framework for these purposes and proposes opportunities to be guided through data analysis tasks, also involved data preparation and assessment. This paper will cover the differences and challenges to integrate Data Mining frameworks like WEKA ([4] and [5]), RapidMiner [6] or R [7] in the sense of [3].

The integrative usage of DSS in SHM is a major challenge concerning "Black Box" systems like BRIMOS$^{\textregistered}$, where integration is not possible analogous to Open Source APIs like WEKA or RapidMiner. Hence, the implementation of a Passive Decision Support Integration System (PDSIS), as a proof of concept for [3], started by integrating operations from RapidMiner [6].

In addition to RapidMiner, WEKA and R, which are briefly described below, there are several problematic issues foreseeable for the integration approach in [3]. Therefore challenges regarding heterogeneities and differences which concern specific notification mechanisms (Server-Client-Application, Web Services, APIs and Libraries, etc.) have to be covered and solved in a semantic integrated manner.

The rest of this paper is structured as follows. Section 2 gives an overview of related work in the domain of integration systems in general, like Observer, DOME or COIN (see [8]). Section 3 covers measurement assessment systems for SHM and examines their functionalities and capabilities of overall decision support in an integrated manner. Section 4 describes PDSIS concepts and their advanced usage of state-of-the-art data mining frameworks like Rapid-Miner or WEKA. Section 5 presents some illustrative examples and compares the results obtained before and after applying additional data mining methods to the transformed output data of BRIMOS$^{\textregistered}$. Finally, in section 6, we summarize the paper and discuss future issues and outstanding scientific questions.

RELATED RESEARCH

This section will give a short overview of appropriate related work of integration systems in general. Approaches like OBSERVER, DOME or COIN are introduced. [9] and [10] compare and describe different approaches of information integration systems in detail and focus on the use of ontologies for integration purposes. In the following we will analyse whether these systems are suitable for the integration of heterogeneous decision support systems in a well-defined workflow [3], [9], [10]:

- SIMS (Search In Multiple Sources): "SIMS appeared in 1993 as an information mediator. The system essentially provides access and integration to multiple sources of information [10]."
- OBSERVER (Ontology Based System Enhanced with Relationship for Vocabulary hEterogeneity Resolution): "OBSERVER is an approach that proposes managing multiple information sources through ontologies [10]."
- DOME (Domain Ontology Management Environment): DOME is focused on ontology development by using software reverse engineering

techniques [10].

- KRAFT (Knowledge Reuse And Fusion/Transformation): "KRAFT was primarily conceived to support configuration design of applications among multiple organizations with heterogeneous knowledge and data models. It uses the concept of "Knowledge fusion" to denote the combination of knowledge from different sources in a dynamic way [10]."
- Carnot: "The Carnot Project was initiated in 1990 with the goal of addressing the problem of logically unifying physically distributed, heterogeneous information [10]."
- InfoSleuth: "The InfoSleuth project is an extension of Carnot to make legacy database systems easily accessible via Web [10]."
- COIN (Context Interchange): "The COIN Project was initiated in 1991 with the goal of achieving semantics interoperability among heterogeneous information sources [10]."

Having a brief introduction of these integration approaches and after studying them in detail [10] one can see that these approaches are mainly used to integrate information and data in ontologies or to combine different information sources. The difference to our integration approach is that besides combining data sources, we also considered the enhancement of collaborative usage of heterogeneous decision support systems based on a predefined general workflow [3]. For more detailed information about the integration approach of heterogeneous decision support systems for Structural Health Monitoring see [3].

MEASUREMENT ASSESSMENT SYSTEMS FOR STRUCTURAL HEALTH MONITORING

Recent research in SHM has been focusing on the development of robust and cost-effective structural health monitoring systems by integrating and extending technologies from various engineering and information technology disciplines.

The *BRIMOS*® (Bridge Monitoring System) [2] technology is based on Ambient Vibration Monitoring and has been used for many years in the field of Structural Health Monitoring. Ambient Vibration Monitoring can be defined as a method for system identification and damage detection in bridge structures due to its dynamic response to ambient excitation such as wind, traffic and micro seismic activity.

BRIMOS® offers a well-defined rating system for investigated structures. This classification allows a fast identification on the structure's integrity as well as the corresponding risk level based on modal parameters, visual inspection reference data.

The rating is a classification based on several research projects, where more than 1000 structures have been assessed and the learned experience has been integrated into the assessment procedure. This classification allows a fast identification of the structure's integrity and the risk level according to the measured dynamic parameters.

SHMTools[1] is a Matlab package that enables the construction of structural

---

[1] http://institute.lanl.gov/ei/software-and-data/SHMTools/

health monitoring (SHM) processes. It provides a set of functions which are organized into modules according to the three primary stages of Structural Health Monitoring: *data acquisition*, *feature extraction* and *feature classification*. It has a modular function design and uses standardized parameter formats which make it easy to create and test various customized SHM processes.

Another SHM system for assessing the health of large structures is introduced in [11]. It uses a networked sensor system and the Matlab Web Server. The system includes an automatic sensor system that transfers the data from the sensor through the Internet and it provides standard evaluation tools to locate and quantify the damage.

PASSIVE DECISION SUPPORT INTEGRATION SYSTEMS

For the purpose of identifying, localizing and assessing damages on structures or structural elements, various different and highly heterogeneous DSS are subject to assist in this process, which includes multiple diverging process steps, each having different syntactical and semantical requirements and communication interfaces to deal with. In order to enforce the integration of multiple heterogeneous DSS, the main concern is the establishment of an integrative system, which aims to implement an additional high-level layer being abstract enough to offer the possibility of covering all conceivable kinds of different DSS in terms of SHM.

Resulting from intensive research within the frame of the European Union IRIS project[2], an integration system *PDSIS* (Passive Decision Support Integration System) has been developed which tends to cover all of the previously mentioned requirements by heavily utilizing ontologies.

Since the participating partners of the IRIS project originate from a broad range of industry sectors, the knowledge and expertise of each sector is constituted in a particular domain ontology. These domain ontologies are very specific and strongly related to the individual industry segments. In order to create an associational and semantical mapping between these, each domain ontology also provides a specific integration ontology which, in turn, uses one central abstract ontology (which we call IRIS ontology) for referring to relevant background information and therefore providing a rather holistic integration approach based on semantic technologies.

The PDSIS architecture consists of a single central user-friendly frontend, which interferes with a mediator component which mainly cares about reasoning the ontologies and extracting the computed individuals and their class assertions [3]. Therefore, for each process step in question, the mediator component is able to suggest suitable systems for a specific task.

Aside from the already mentioned BRIMOS® system, numerous data mining frameworks are used for decision support and have been integrated within PDSIS. Among others, these frameworks are:

---

[2]Integrated European Industrial Risk Reduction System (IRIS), Funded under 7th FWP (Seventh Framework Programme), Research area: NMP-2007-3.1-3. See `http://www.vce.at/iris/`

RapidMiner A data mining toolkit, well-suited for processing very large data pools and performing various analytical tasks. Being open source, it offers a programming interface and is primarily intended for data mining [6] operations which makes it a considerable utility for decision support. At the time of this writing, the mediator service fully implements the RapidMiner *Fast Fourier Transformation* and *Principal Component Analysis* functionalities.

WEKA: Originating from the University of Waikato, WEKA (*Waikato Environment for Knowledge Analysis*) [5] aims to be a versatile framework for machine learning and predictive modeling including but not limited to preprocessing, text-learning and clustering. Currently, the corresponding IRIS mediator implements *Wavelet* transformation using the standard WEKA Java API. *Wavelet Transformation.* The specific notification workflow can be interpreted as analogous to RapidMiner.

R (GNU S): As opposed to the two frameworks outlined above, R [7] is a non-graphical programming language for statistical computations and became a de facto standard for statistical software development. R implements the S programming language and originates from the University of Auckland. Since R is not developed in Java, it does not offer an integrated programming interfaces like RapidMiner and WEKA, although certain third-party expression evaluator interfaces exist [12]. Currently, R has not yet been implemented as mediator service within PDSIS, but approaches towards including R through either web-services or socket communication are in progress. Furthermore, R also supports graphical plotting which, at this time, is not intended to be utilized by PDSIS.

CASE STUDY

In the following sections we will provide and analyze an example of applying different operations from the already described Data Mining and Machine Learning Frameworks to the measurement analysis process. It shall be pointed out explicitly that the below operations, which were applied to the analysis process, are just examples of how to integrate these operations with tools and applications from the SHM domain. Definitely, the intention of this article it is not to show *which* operations add the best value to measurement assessment, but to prove that by applying Machine Learning techniques it can increase the runtime performance of calculations, increase precision or understandability of results or offer the possibility to gain completely new knowledge about a measurement.

The S101 is a concrete bridge which was built in the 1960ies. Because it required maintenance and was too small for the location, it was decided to be replaced. This offered a great opportunity to perform a damage test to demonstrate the impact of scientific insight and findings with regard to reduced uncertainties. The results are meant to support the decision process of infrastructure owners in the course of cost planning for maintenance and possible rehabilitation measures in general.

Calculation of Eigenfrequencies with BRIMOS®

BRIMOS® is a sophisticated tool for measurement analysis, which was developed and is operated by the Vienna Consulting Engineers[3]. The following Figure 1 shows the Eigenfrequencies, which were calculated from a particular measurement (taken from the bridge "S101").

The below measurement consisting of 48 channels (16 sensors measuring acceleration in 3 dimensions each) further on was Input for Rapidminer's Principal Components Analysis (PCA), which was re-analyzed using BRIMOS®.
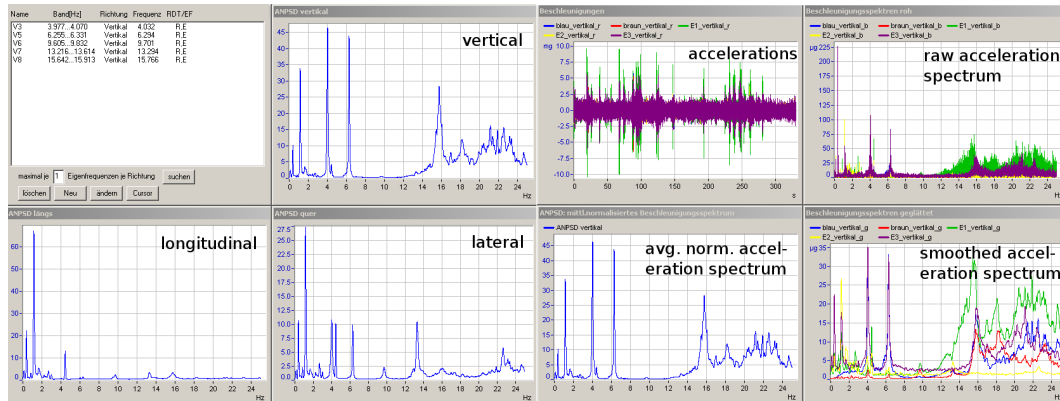

Figure 1: Eigenfrequencies from Original Measurement Data (165000 Measure Points, 48 Channels)

Figure 2 shows the results from measurements of full length, but with a reduced number of dimensions (principal components).

Comparing Figure 2 with Figure 1, one can see that mainly the heterodyne or redundant components of the signal were abstracted.

The signal still contains the main information. This visual assessment is underlined by the results shown in table I.

For the calculation of Eigenfrequencies a reduction of measurement channels using PCA is a reasonable operation (see table I), but, there are some performance indicators, which depend on sensor positions or acceleration directions (e.g. mode shapes). These Figures cannot be calculated (at least not directly) from a measurement which was reduced by PCA.
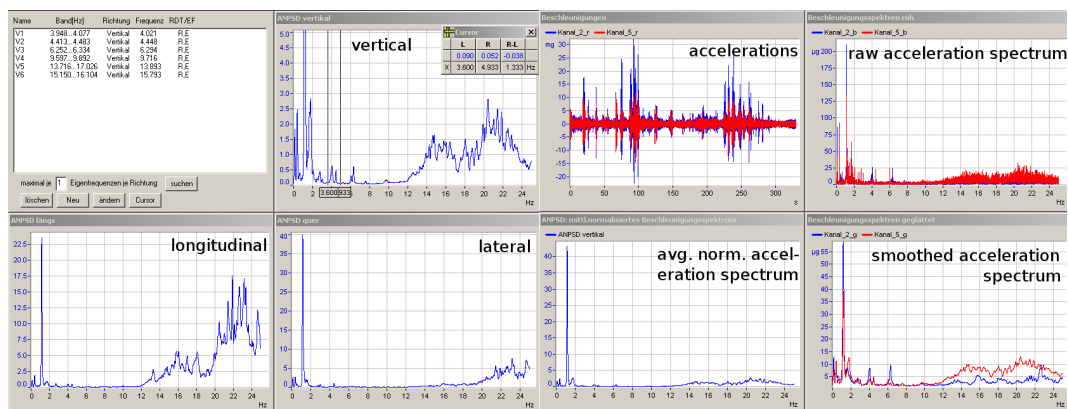

Figure 2: Eigenfrequencies after Calculation of Principal Components (165000 random Measure Points, 6 Channels)

---

[3]http://www.vce.at

TABLE I: QUANTITATIVE RESULTS

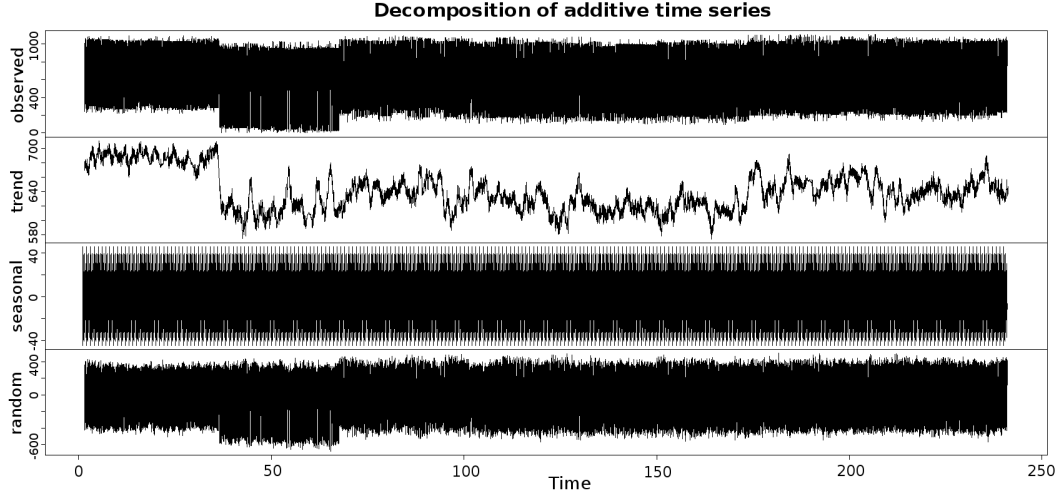| Eigen-frequency	Original, full length	14 Channels, full length	6 Channels, full length	21 Channels, reduced length
EF1	4.032	4.032	4.021	4.059
EF2	6.294	6.294	6.294	6.287
EF3	9.701	9.701	9.716	9.705
EF4	13.294	13.294	13.893	13.275
EF5	15.766	15.766	15.793	15.747

Applying filters from R

Figure 3: Decomposing the eigenfrequencies time series into trend, seasonal and random components

The above plot was obtained by filtering the time series of Eigenfrequencies using the *decompose* (additive) function from R. Thus, we obtained a decomposition of the Eigenfrequencies into trend, seasonal and random or residual components. In general, time series decompostion allows us to distinguish differences in the data due to the actual trend, seasonal variation and error components. It will also allow us to predict future values based on the trend and seasonal effects.

Comparison

The above examples show how various data mining methods can be used at different stages of the data analysis process in the frame of the SHM domain. While using only BRIMOS® we can obtain only a transformation of the raw data, by applying PCA after this transformation we can reduce significantly the dataset. This way we keep the relevant information while enabling a faster and easier further analysis using a smaller dataset. Also, by applying several other filters from R (e.g. decomposition, linear filters) we can get a better insight of how the data behaves in time and which components are seasonal.

CONCLUSION AND FURTHER WORK

This paper describes general procedures and concrete implementations of Data Mining operations from RapidMiner and WEKA to be used by tech-

nicians via PDSIS. Therefore a mediator component uses the semantic descriptions of the integrated operators and the semantic requirements of the different process step types (both topics in [3]) to provide lists of available and executable functions. The current version of PDSIS, as a web application for end user interactions, and IRISServices, as service pool and/or provider of different web services. Several prospective improvements have to be covered:

- Integration of operators for assessment tasks and combination with data preparation part
- First draft for complete workflow execution
- Validation of operators
- Comparison of different operator results
- Changes of operator parameters should be sent to mediator and used for further execution
- Workflow planning tool, tob be included in PDSIS web application and provide functionalities to create a user-defined execution plan.
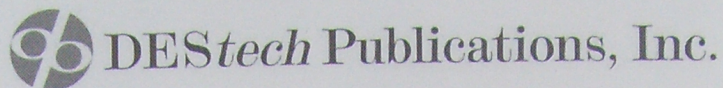
REFERENCES

1. Wenzel, H., Pichler, D.: Ambient Vibration Monitoring. John Wiley & Sons Ltd., Chichester, 2005
2. Wenzel, H.: Health Monitoring of Bridges. John Wiley & Sons Ltd., Chichester, 2008
3. Anderlik, S., Stumptner, R., Freudenthaler, B., Fritz, M.: A Proposal for Ontology-based Integration of Heterogeneous Decision Support Systems for Structural Health Monitoring. In Proceedings of the 12th International Conference on Information Integration and Web-based Applications and Services, pp. 166–173, Paris, 2010
4. Graczyk, M., Lasota, T., Trawinski, B.: Comparative Analysis of Premises Valuation Models Using KEEL, RapidMiner, and WEKA. In Proceedings of 1st International Conference on Computational Collective Intelligence - Semantic Web, Social Networks & Multiagent Systems, pp. 800–812, Berlin/Heidelberg, 2009
5. Holmes, G., Donkin, A., Witten, I.H.: WEKA: a machine learning workbench. In Proceedings of the 1994 Second Australian and New Zealand Conference on Intelligent Information Systems, pp. 357–361, Brisbane, Qld., 1994
6. Rapid-I: RapidMiner 5.0 Manual. http://rapid-i.com/content/view/26/84/, 2010
7. Jones, O., Maillardet, R., Robinson, A.: Introduction to Scientific Programming and Simulation Using R. Chapman & Hall/CRC, 2009
8. Wache, H., Voegele, T., Visser, U., Stuckenschmidt, H., Schuster, G., Neumann, H. and Huebner, S.: Ontology-Based Integration of Information - A Survey of Existing Approaches. In Proceedings of the IJCAI-01 Workshop on Ontologies and Information Sharing (Seattle, USA, August 4 - 5, 2001), IJCA'01, pp. 108–117, IEEE, 2001
9. Buccella, A., Cechich, A. and Brisaboa, N.R. 2005. Ontology-Based Data Integration Methods: A Framework for Comparison. Paper Proposal. University of Comahue.sources. Journal of Intelligent and Cooperative Information Systems 2, 2, 127-158.
10. Wache, H., Voegele, T., Visser, U., Stuckenschmidt, H., Schuster, G., Neumann, H. and Huebner, S. (2001). Ontology-Based Integration of Information - A Survey of Existing Approaches. In Proceedings of the IJCAI-01 Workshop on Ontologies and Information Sharing, IEEE, pp. 108–117
11. Mita, A., Inamura, T., Yoshikawa, S.: Structural Health Monitoring for Buildings with Automatic Data Management System. In Proceedings of the 4th International Conference on Earthquake Engineering, Taiwan, 2006.
12. Ripley B. D.: The R and Omegahat projects in statistical computing. Department of Statistics, University of Oxford, 2001

---

[4]http://cordis.europa.eu/fp7/

DEStech Publications, Inc.

CONTRIBUTING AUTHOR COPYRIGHT RELEASE FORM

As author of the chapter/contribution titled _INTEGRATION OF DATA MINING OPERATIONS FOR STRUCTURAL HEALTH MONITORING_

appear in the *Proceedings of Structural Health Monitoring 2011*, I hereby agree to the following:

1. To grant to DEStech Publications, Inc., 439 North Duke Street, Lancaster, PA, 17602, copyright of the above named chapter/contribution (for U.S. Government employees to the extent transferable), in print, electronic, and online formats. However, the undersigned reserve the following:
a. All proprietary rights other than copyright, such as patent rights.
b. The right to use all or part of this article in future works.

   DEStech Publications thereby retains full and exclusive right to publish, market, and sell this material in any and all editions, in the English language or otherwise.

1. I warrant to DEStech Publications, Inc., that I am the (an) author of the above-named chapter/contribution and that I am the (a) copyright holder of the above-named chapter/contribution granted to DEStech Publications, Inc.

2. I warrant that, where necessary and required, I have obtained written permission for the use of any and all copyrighted materials used in the above-named chapter/contribution. I understand that I am responsible for all costs of gaining written permission for use of copyrighted materials.

3. I agree to assume full liability to DEStech Publications, Inc. and its licensee, and to hold DEStech Publications, Inc. harmless for any claim or suit filed against DEStech Publications, Inc. for violation of copyrighted material used in the above-named contribution.

Please sign and date this form and retain a copy for your records. Please include original form with your chapter/paper.

Thank you for your cooperation.

Please print name: _ERIK SONNLEITNER_

Signed: _____ Dated: _2011-05-14_