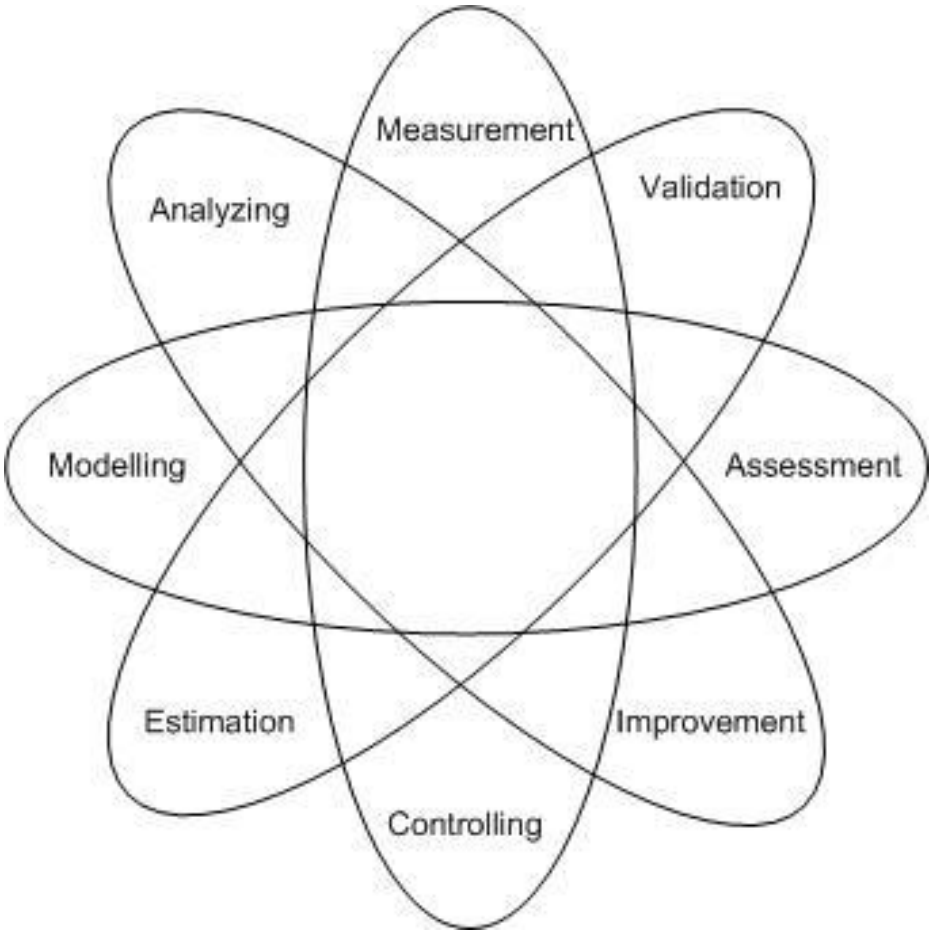
DASMA

Volume 23, Number 1, March 2018

Software Measurement News

Journal of the Software Metrics Community



Editors:

Alain Abran, Manfred Seufert, Reiner Dumke, Christof Ebert, Cornelius Wille

Université du Québec
École de technologie supérieure

COSMIC

SM L@b

CONTENTS

Announcements **2**

Conference Reports **7**

News Papers **17**

Abdalla Bala, Alain Abran
Impact Analysis of Multiple Imputation on Effort Estimation Models with the ISBSG Repository of Software Projects 17

Andreas Schmietendorf, André Nitze
Empirische Untersuchung zum Einsatz von SaaS-Lösungen bei Klein- und mittelständischen Unternehmen (KMUs) 35

New Books on Software Measurement **49**

Conferences Addressing Measurement Issues **55**

Metrics in the World-Wide Web **63**

Editors:

Alain Abran
Professor and Director of the Research Lab. in Software Engineering Management
École de Technologie Supérieure - ETS, 1100 Notre-Dame Quest,Montréal, Quebec, H3C 1K3, Canada, ***Tel.:*** *+1-514-396-8632,* ***Fax:*** *+1-514-396-8684*
alain.abran@etsmtl.ca

Manfred Seufert
Chair of the DASMA, MediaanABS Deutschland GmbH
Franz-Rennefeld-Weg 2, D-40472 Düsseldorf, ***Tel.****: +49 211 250 510 0*
manfred.seufert@mediaan.com

Reiner Dumke
Professor on Software Engineering, University of Magdeburg, FIN/IKS
Postfach 4120, D-39016 Magdeburg, Germany,
dumke@ivs.cs.uni-magdeburg.de, http://www.smlab.de

Christof Ebert
Dr.-Ing. in Computer Science, Vector Consulting Services GmbH
Ingersheimer Str. 20, D-70499 Stuttgart, Germany, ***Tel.:*** *+49-711-80670-1525*
christof.ebert@vector.com

Cornelius Wille
Professor on Software Engineering, University of Applied Sciences Bingen
Berlinstr. 109, D-55411 Bingen am Rhein, Germany,
Tel.: *+49-6721-409-257,* ***Fax:*** *+49-6721-409-158*
wille@fh-bingen.de

Editorial Office: University of Magdeburg, FIN, Postfach 4120, 39016 Magdeburg, Germany

Technical Editor: Dagmar Dörge

The journal is published in one volume per year consisting of two numbers. All rights reserved (including those of translation into foreign languages). No part of this issues may be reproduced in any form, by photoprint, microfilm or any other means, nor transmitted or translated into a machine language, without written permission from the publisher.

© 2018 by Otto-von-Guericke-University of Magdeburg. Printed in Germany

IWSM-MENSURA September 19-20, 2018, Beijing

Venue: Beijing National Convention Center
http://www.iwsm-mensura.org/

Call for participation: Plenary Talks (1 hour), Short Talks (30 minutes), Tutorials

We solicit proposals from experienced industry practitioners and researchers with COSMIC expertise and experience in the usage of COSMIC. The main goal is to share experiences, challenges and solution approaches to facilitate technology transfers from best estimation-related practices developed by researchers and world industry experts.

THEME & SCOPE: COSMIC Function Points: Fundamentals for Software Effort Estimation

- COSMIC usage in industry across the world, and the harvesting of the related benefits.
- COSMIC Measurement best practices
- State of the art on COSMIC research achievements & technology transfers
- Estimation improvements through COSMIC usage
- Automation of COSMIC from source code
- Automation of COSMIC from requirements
- Earned-value management with COSMIC
- How to implement COSMIC in Agile environments
- Software Measurements Programs: Industry Experience report
- Decision support systems based on software measurement
- Visualizations and dashboards

Software and IT measurement are keys for successfully managing and estimating software development projects: sound measurement techniques, such as COSMIC Function Points, are essential for business and engineering: they enrich technical knowledge regarding the practice of software development estimation, as well as empirical research in software estimation.

Submissions: Proposals of Plenary Talks, Short Talks and Tutorial must include a Title, a max of 1 page and a short bio, and must be submitted to one of the program co-chairs.

At least one among the authors of each Talk submission accepted must register for the conference and commit to their presentation. Accepted & presented talks (Papers and-or Powerpoints) will be posted on the conference website for post-conference public access.

IMPORTANT DATES

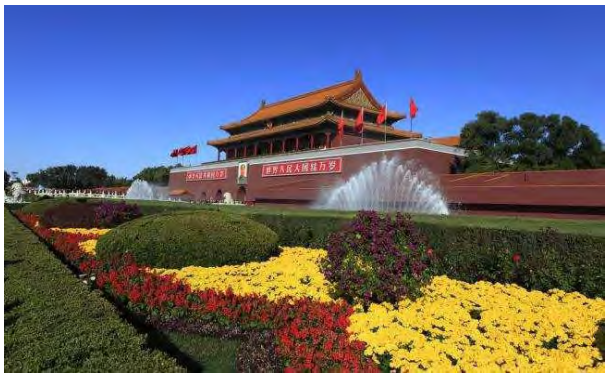
Plenary Talks – Short Talks -Tutorial submissions before : **May 1st, 2018**
Notification of acceptance/rejection: **June 1, 2018**
Final version: **September 1, 2018**

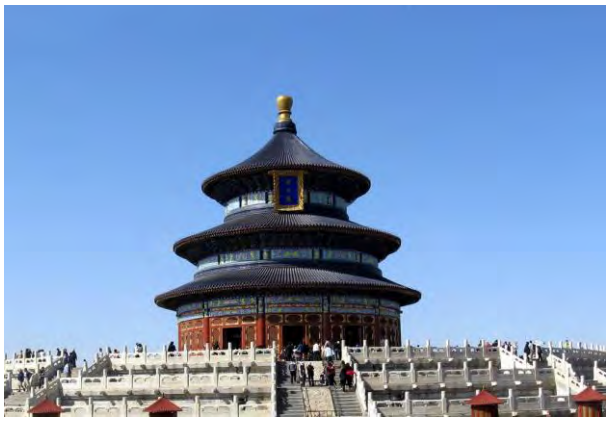
CONTACT INFORMATION - Program co-Chairs
Dylan Ren: renjialin@measures.net.cn
Sylvie Trudel: trudel.s@uqam.ca

Important notes:

- Participants to the IWSM-MENSURA are invited to attend the Chinese national 'Cost Measurement Conference' on September 18, at the same venue.
- SPI-China will host a post-conference 1-day tour of Beijing (free activity).

Impressions of the location:





Sponsoring:

SPI China
系统与软件过程改进

COSMIC

Measures 麦哲思科技

Agile 上海艾纵管理咨询

(to be added)

API-First/API-Management

Open APIs als Treiber der Digitalisierung

ECC, 18.04.2018 (09:00 bis 17:00 Uhr), Hamburg

Moderne Applikationen machen umfangreichen Gebrauch von offenen und webbasiert zur Verfügung gestellten APIs. Typische die Digitalisierung treibende Lösungen finden sich mit mobilen Applikationen (Apps), Anwendungen der künstlichen Intelligenz (Data Science), im Telematikbereich (Smart City) oder auch beim Internet der Dinge (Smart Home). Wollen Unternehmen den digitalen Wandel erfolgreich gestalten, gilt es API-Ansätze im eigenen Unternehmen zu identifizieren, aber auch den Einsatz extern verfügbarer Open APIs in Betracht zu ziehen. Das Schnittstellendesign und das Management über den gesamten Lebenszyklus einer Open API bestimmen dabei maßgeblich über Erfolg und Misserfolg.

Keynote im Rahmen der ECC-Tagung (09:15 Uhr)

Rüdiger Wölfl (CISCO Systems GmbH)

Keynote: Digitalisierung

Start des Workshops

Session 1 – Eröffnung und Einführung (10:30 Uhr):

Prof. Dr. Andreas Schmietendorf (HWR Berlin)

API-Fizierung als Treiber der Digitalisierung

- Motivation für Open Data und Open API
- API-Management im Life Cycle

André Nitze (UltraTendency)

Contract-first Development

- Service-Design für nachhaltige API-Angebote
- OpenAPI/Swagger-basierte Verträge

12:00 bis 13:30 Mittagspause

Session 3 – Impulsvorträge & Diskussion (13:30 bis 15:00 Uhr):

Sandro Hartenstein (Hartenstein IT Security Services)

Bridging the gap - Enterprise Security gestern und heute

- Identity and Access Management (IAM)
- Open Source Enterprise-Lösung "Keycloak"

Tobias Kiertscher (TH Brandenburg)

Docker ohne Cloud - Ein Erfahrungsbericht

- Softwaredefinierte Infrastrukturen

- DevOps basierte Servicebereitstellung

15:00 bis 15:30 Kaffeepause

Session 4 – Hands On Seminar (15:30 bis 17:00 Uhr):

Lukas Böhme, Jonas Grunert (Bayer AG/HWR Berlin)

Abfragen auf HTTP-basierte Service-APIs

- Möglichkeiten zur selektiven Abfrage
- Facebooks GraphQL im Praxistest

Abschluss der Veranstaltung

Die korrespondierenden Vorträge der Referenten werden den Teilnehmern über die Webseite der ceCMG (www.cecmg.de) zur Verfügung gestellt. Die Ergebnisse der entsprechenden Diskussionsrunden werden zeitnah im Internet publiziert. Änderungen am Programm sind unter Vorbehalt möglich. Für Verpflegung vor Ort wird gesorgt.

Für die Teilnahme an der Veranstaltung ist eine kostenpflichtige Anmeldung zur Enterprise Computing Conference (ECC 2018) erforderlich. Für Mitglieder der ceCMG-, DASMA-, GI- und ASQF gilt eine reduzierte Teilnahmegebühr. Über die Teilnahmegebühr erhalten Sie eine Rechnung der ceCMG e.V. (Central Europe Computer Measurement Group).

Veranstaltungsort: *Hamburg, Lindner Park-Hotel Hagenbeck*

Weiteren Informationen und Anmeldung unter: http://www.cecmg.de

Kontakt: Susanne Mund – sekretariat@cecmg.de

Workshop
"Evaluation of Service-APIs – ESAPI 2018"

Munich, November 2018

see: https://blog.hwr-berlin.de/schmietendorf

GI FG Veranstaltungen

Data Science ist ein neuer und schnell wachsender Schwerpunkt in Software Messung.

Unser Ziel: Grundlagen verstehen, Methodik kennen, und aus praktischen Anwendungsfälle lernen. Daher ist die Jahres-veranstaltung 2018 der GI FG Software Measurement zum Thema "Data Science" im Rahmen der GI-Jahreskonferenz vom **26. - 27. September.**

Informatik 2018 – GI-Jahrestagung in Berlin

IWSM/Mensura 2017

25.-27.10.2017 Gothenburg, Sweden

Miroslaw Staron, Wilhelm Meding: *Joined Conference of the 27th International Workshop on Software Measurement (IWSM) and the 12th International Conference on Software Process and Product Measurement (Mensura), ACM 2017, ISBN 978-1-4503-4853-9*

Proceedings Contents

(adapted from the university bibliography service http://dblp.uni-trier.de/db/conf/iwsm/iwsm2017)

Agile/Distributed Software Development

Using measurement and simulation for understanding distributed development processes in the cloud
Lunescu, I.; Marchesi, M.; Münch, J.; Kuhrmann, M. *(pp. 1 - 11)*

Identifying and mitigating risks of software project management in global software development
Chadli, S. Y.; Idri, A. *(pp. 12 - 22)*

Effective monitoring of progress of agile software development teams in modern software companies: an industrial case study
Meding, W. *(pp. 23 - 32)*

Agile and lean metrics associated with requirements engineering
Wnuk, K.; Maddila, K. C. *(pp. 33 - 40)*

Effort estimation for agile software development: comparative case studies using COSMIC functional size measurement and story points
Salmanoglu, M.; Hacaloglu, T.; Demirörs, O. *(pp. 41 - 49)*

Adopting flow analysis in software development projects
Estevam, A.; Dennehy, D.; Conboy, K. *(pp. 50 - 53)*

Software Quality

Autonomous real-time software & systems testing
Fehlmann, T.; Kranich, E. *(pp. 54 - 63)*

Using FSM patterns to size security non-functional requirements with COSMIC
Ungan, E.; Trudel, S.; Poulin, L. *(pp. 64 - 76)*

A shortcut to estimating non-functional requirements?: architecture driven estimation as the key to good cost prediction
Vogelezang, F. W.; Poort, E. R.; Vliet, E.v.d.; Mols, H.R.J.; Nijland, R. (pp. 77 - 81)

Filling the gaps: imputation of missing metrics' values in a software quality model
Kupinski, S.; Walter, B.; Wolski, M.; Chojnacki, J. *(pp. 82 - 87)*

Discovering maintainability changes in large software systems
Molnar, A.; Motogna, S. *(pp. 88 - 93)*

Measuring Process Planning

Sustainable measurement programs for software development companies: what to measure
Meding, W. *(pp. 94 - 99)*

Improving the real-time experience for software measurement system end-users
Wang, H.; Hebig, R. *(pp. 100 - 106)*

Automation

Automatically measuring the maintainability of service- and microservice-based systems: a literature review
Bogner, J.; Wagner, S.; Zimmermann, A. *(pp. 107 - 115)*

Automated analyses of model-driven artifacts: obtaining insights into industrial application of MDE
Mengrink, J.; Serebrenik, A.; Schiffelers, R.R.H.; Brand, M.G.J.v.d. *(pp. 116 - 121)*

Functional Size Measurement

Functional size measurement for the internet of things (IoT): an example using COSMIC and the arduino open-source platform
Soubra, H.; Abran, A. *(pp. 122 - 128)*

A rapid measurement procedure for sizing web and mobile applications based on COSMIC FSM method
Haoues, M.; Sellami, A.; Ben-Abdallah, H. *(pp. 129 - 137)*

Towards an automated functional size measurement procedure: an industrial case study
Quesada-Lopez, C.; Jenkins, M.; Salas, L. C.; Gomez, J. C. *(pp. 138 - 144)*

Project Management

Earned scope management: a case study of scope performance using use cases as scope in a real project
Souto, F. V. *(pp. 145 - 160)*

A new approach for continuously monitoring project deadlines in software development
Fehlmann, T. M.; Kranich, E. *(pp. 161 - 169)*

Using weekly open defect reports as an indicator for software process efficiency: theoretical framework and a longitudinal automotive industrial case study
Mellegard, N. *(pp. 170 - 175)*

Mining social networks of open source CVE coordination
Ruohonen, J.; Hyrynsalmi, S.; Rauti, S.; Leppanen, V. *(pp. 176 - 188)*

3-D visualization of dynamic runtime structures
Ciolkowski, M.; Faver, S.; Mammen, S. v. *(pp. 189 - 198)*

Effort Estimation

Effort estimation methods for ERP projects based on function points: a case study
Ömüral, N. K.; Demirörs, O. *(pp. 199 - 206)*

Investigating heterogeneous ensembles with filter feature selection for software effort estimation
Hosni, M.; Idri, A.; Abran, A. *(pp. 207 - 220)*

Software estimation: towards prescriptive analysis
Pospieszny, P. *(pp. 221 - 226)*

COSMIC (ISO 19761) certification: analysis of examinee results
Desharnais, J.M.; Abran, A. *(pp. 227 - 231)*

Software Design and Coding

Measurement of enterprise architecture (EA) from an IT perspective: research gaps and measurement avenues
Ilin, I. V.; Levina, A. I.; Abran; A.; Iliashenko, O. J. *(pp. 232 - 243)*

Improving object-oriented design quality: a portfolio- and measurement-based approach
Bräuer, J.; Saft, M.; Plösch, R.; Körner, C. *(pp. 244 - 254)*

Correlations of software code metrics: an empirical study
Mamun, A. A.; Berger, C.; Hansson, J. *(pp. 255 - 266)*

for more information, see: http://www.iwsm-mensura.org/

and http://dblp.uni-trier.de/db/conf/iwsm/iwsm2017

Der Workshop fand wieder unter exzellenten Bedingungen am IESE in Kaiserslautern statt. Die Inhalte der Vorträge sind unter **http://www.software-messung.de/** einsehbar.

Workshop "Evaluation of Service-APIs – ESAPI 2017" Service APIs als Enabler einer erfolgreichen Digitalisierung

(detaillierter Bericht)

Michael Binzen, Konrad Nadobny, Hendrik Neumann, André Nitze, Andreas Schmietendorf

1. Motivation zum Workshop

Im Internet als Service zur Verfügung gestellte Informationen, Funktionen und Algorithmen bestimmen in zunehmendem Maße die Art und Weise, wie neue Softwarelösungen implementiert werden. Im technologischen Sinne handelt es sich dabei um webbasiert zur Verfügung gestellte Daten und Service-APIs, die eine „ad hoc"-Integration in neue Lösungskontexte unterstützen. Entsprechende Lösungen finden sich z.B. mit mobilen Applikationen (Apps), Anwendungen der künstlichen Intelligenz (Data Science), im Telematikbereich (Smart City) oder auch im Internet der Dinge (IoT). Wollen Unternehmen von den Möglichkeiten der Digitalisierung profitieren, ist es erforderlich, die eigenen Bedürfnisse und Möglichkeiten realistisch bewerten zu können. Aus dieser Tatsache ergeben sich mindestens zwei Perspektiven auf Open-Data und Open-APIs: die Nutzersicht und die Entwicklersicht.

Darüber hinaus gilt es den gesamten Lebenszyklus der angebotenen Daten und APIs einem Management zu unterziehen. Ein solches API-Management muss sowohl den Bedürfnissen der Serviceentwicklung, einer heute zumeist agil durchgeführten Serviceintegration, als auch denen des betrieblichen Einsatzes der dann zusammengesetzten Services entsprechen. Typische Managementfunktionen beziehen sich auf den gesamten Lebenszyklus, die Zugriffskontrolle, das Monitoring, die Verrechnung oder auch auf die eingesetzten Laufzeitumgebungen der Service-APIs.

Ein besonderes Spannungsfeld ergibt sich aus offen und kostenfrei zur Verfügung gestellten Schnittstellen (Open Data/Open API) und ausschließlich kommerziell angebotenen Schnittstellen (API economy). Im Zusammenhang mit der API economy finden sich schnell wachsende Unternehmen (z.B. Fahrdienstvermittler Uber, Unterkunftsvermittler Airbnb, Onlineversand Zalando), die zwar einen breiten Marktzugang besitzen, aber nur eine geringe Ressourcenbindung aufweisen. Alleinstellungsmerkmale dieser die Digitalisierung treibenden Unternehmen beziehen sich auf die optimale Erfüllung der Kundenbedürfnisse. Dafür benötigte Leistungen werden von einem globalen Markt unter Verwendung der hier im Mittelpunkt stehenden Service-APIs im Sinne einer agil gebildeten Lieferantenkette bezogen. Darüber hinaus werden unternehmensexterne Softwareentwickler bewusst am Innovationsprozess bisher eher geschlossener Unternehmen beteiligt, so dass mit Hilfe von Service-APIs Möglichkeiten zur Gestaltung globaler einsetzbarer Softwarelösungen entstehen. Dieser Sachverhalt unterstreicht die strategische Bedeutung von derart zur Verfügung gestellten Informationen, Funktionen und Algorithmen.

Die aus dem Diskurs der GI-Fachgruppe „Software-Messung und –Bewertung" hervorgegangene ESAPI-Initiative widmet sich multidimensionalen Bewertungsaspekten der vielfältig im Web angebotenen Service-APIs. Aus technologischer Sicht handelt es sich bei diesen Service-APIs um Web Services, Web-APIs, Microservices und Angebote, die unter dem Begriff Open Data bzw. Open API subsumiert werden. Mit der inhaltlichen Fokussierung wird den Themenschwerpunkten vergangener Veranstaltungen der BSOA/BCloud-Interessensgemeinschaft Rechnung getragen.

2. Beiträge des Workshops

Um der Zielstellung eines interaktiv gestalteten Workshops Rechnung zu tragen, wurden im Vorfeld Keynotes, Impulsvorträge und Kurzseminare durch ein deutschlandweit zusammengesetztes Programmkomitee ausgewählt.
Für die Inhalte der Beiträge sei auf den publizierten Tagungsband verwiesen. Folgende Autoren beteiligten sich daran mit ihren Beiträgen:

Michael Binzen: (Deutsche Bahn AG):

> **Keynote**: Revolution durch Evolution – mit APIs schrittweise zu mehr Business Agility und neuen Geschäftsmodellen

Dr. Frederik Kramer (initOS GmbH)

> Impulsvortrag: Case Study zur Bewertung einer sinnvollen Nutzung von Microservices

Prof. Dr. Olaf Resch (HWR Berlin)

> Impulsvortrag: Die API-Economy braucht ein iUDDI

Konrad Nadobny, Andreas Schmietendorf (Bayer AG, HWR Berlin)

> Impulsvortrag: Empirische Analysen von Open-Data und Open-API „Service“ Angeboten

Steffen Kosterski (Toll Collect)

> Keynote: Compliance-Aspekte bei der Datenerhebung und Verarbeitung

André Nitze (TH Brandenburg, OvG-Universität Magdeburg)

> Kurzseminar: Contract-first Development - Vom OpenAPI (Swagger) basierten Vertrag zur Implementierung

Lukas Böhme, Jonas Grunert (Bayer AG)

> Kurzseminar: Bewertung der Vor- und Nachteile einer GraphQL basierten Abfrage von HTTPbasierten Service-APIs

3. Ergebnisse des World-Cafés

Eine der Zielstellungen des ESAPI-Workshops war, dass die Teilnehmer ihre Standpunkte, Sichtweisen und Ideen zu ausgewählten Fragen rund um das Thema Open API bzw. Open Data austauschen. Hierfür wurde die interaktive Methode eines World-Cafés gewählt, da es den Teilnehmern auf diese Weise ermöglicht wurde, in begrenzter Zeit ohne vorherige Vorbereitung die Themen strukturiert zu diskutieren.

3.1 Impulsfragen zur Abstimmung

Im Vorfeld des Workshops wurden den Teilnehmern die folgenden Fragen zwecks Priorisierung und ggf. Feedback zur Verfügung gestellt:

1. Welche Bedeutung hat Open Data bzw. Open API im Kontext globaler Unternehmensarchitekturen (Platform-Ecosystems)?

2. *Wie kann Open Data bzw. Open API zu Open Innovation führen und damit zum Wettbewerbsfaktor werden?*

3. *Wie kann Open Data bzw. Open API in eine umfassende Digitalisierungsstrategie eingebettet werden?*

4. Wie ist das Verhältnis von Open Data bzw. Open APIs zu den Rahmenbedingungen einer API-economy?

5. Wie können die Inhalte des Open Data bzw. Open API Manifests der BITKOM für das eigene Unternehmen operationalisiert werden?

6. Welche Einflüsse gilt es beim Management von Open Data und Open API Serviceangeboten zu berücksichtigen?

7. *Welche Aspekte gilt es beim Sicherheits- und Compliance-Management zu berücksichtigen, was ist aktuell „state-of-the-art"?*

Durch die Teilnehmer wurden die Fragen 2, 3 und 7 für eine vertiefende Diskussion innerhalb der World-Cafés ausgewählt. Dies erfolgte per Mehrheitsabstimmung, wobei jeder Workshop-Teilnehmer vier Stimmen auf die vorgeschlagenen Themen frei vergeben konnte. Im Zusammenhang mit ausgewählten Fragen wurden erste Aspekte des korrespondierenden Leitfadens der BITKOM (vgl. [BITKOM 2017]) diskutiert.

3.2 Ergebnisse der Diskussion

Wie kann Open Data bzw. Open API in eine umfassende Digitalisierungsstrategie eingebettet werden?

Open Data bzw. Open API ist als Instrument zur Umsetzung einer strategischen Entscheidung zur Öffnung des Unternehmens zu verstehen und somit ein wichtiger Enabler im Rahmen einer umfassenden Digitalisierungsstrategie. Damit Open Data/Open API erfolgreich für das Erreichen strategischer Ziele eingesetzt werden kann, müssen einige Faktoren in Betracht gezogen werden. Die Entscheidung für die Umsetzung von Open Data/Open API sollte zunächst auf Grundlage einer Nutzwertanalyse getroffen werden, wobei Risiken und Benefits auch unter Betrachtung von Case Studies gefunden und abgewogen werden sollten. Ausgehend von einer Analyse des Status-Quo (interne und externe Rahmenbedingungen) müssen dann Wege gefunden werden, wie Open Data/Open API strategisch umgesetzt werden kann.

Versteht man unter einer Strategie die Regeln und Maßgaben für das unternehmerische Handeln, so sind zunächst die Etablierung einer guten Governance und die Schaffung der Voraussetzungen wichtig. Das Bekenntnis zu Open Data/Open API steht am Anfang des Transformationsprozesses. Es muss ein Rahmenwerk etabliert werden, welches dem einzelnen Mitarbeiter die Sicherheit gibt, dass sein Handeln hin zu Open Data/Open API richtig und erwünscht ist. Dieses Rahmenwerk muss hierbei flexibel genug sein, dass es innovative Lösungen erlaubt und dabei gleichzeitig klar genug ist, um Unsicherheit zu vermeiden. Auch die organisatorische Struktur muss an die entsprechende Strategie angepasst werden. Ein Treiber für die Umsetzung von Open Data/Open API sind dabei die jeweiligen Business-Needs, welche die technischen Lösungen bestimmen. Diese zu erkennen, zu fördern und Erfolgsgeschichten zu kommunizieren ist essentiell für den Erfolg der Strategie. Nur durch die Schaffung der richtigen Strukturen, eine konstante Kommunikation und Reflektion sowie eine stringente Umsetzung im operativen Alltag kann der kulturelle und organisatorische Wandel erfolgreich umgesetzt werden, welcher für die Implementierung einer Open Data/Open API-Strategie erforderlich ist.

Zusammenfassend kann Open API/Open Data in eine umfassende Digitalisierungsstrategie eingebettet werden, indem die richtigen Rahmenbedingungen geschaffen werden und die Transformation in Einklang mit den Unternehmenszielen steht. Open Data/Open API kann somit zu einem entscheidenden Treiber für die digitale Transformation und die Innovationskraft eines Unternehmens werden.

Welche Aspekte gilt es beim Sicherheits- und Compliance-Management zu berücksichtigen, was ist aktuell „state-of-the-art"?

Bei der Implementierung von Open Data/Open API stellt sich zwangsläufig die Frage, welche Sicherheitsrisiken den Vorteilen gegenüberstehen. Als Vorteile von Open Data/Open API gelten vor allem das freigesetzte Innovationspotential sowie die Öffnung für Synergien, neue Geschäftsmodelle, die Demokratisierung der Daten und Open Data/Open API als Enabler für neue Lösungen. Dem gegenüber stehen Risiken in Bezug auf Datenschutz und Datensicherheit sowie der allgemeinen Sicherheit der über Open API angebotenen Services. Diese Risiken sind beim Sicherheits- und Compliance-Management zu berücksichtigen.

Der freie Zugriff auf Daten und Services über Open Data/Open API eröffnet zunächst das Risiko des kriminellen Missbrauchs, gezielter Angriffe auf die Services sowie der Ownership und Haftung für Daten und Services auf der Anbieterseite. Auf der Nachfrageseite besteht das Risiko der Fehlinterpretation der Daten und Unsicherheit bezüglich richtiger Funktion, Verlässlichkeit und Verfügbarkeit der nachgefragten Services. Zudem sollte der Service einfach und unkompliziert eingebunden werden können.

Den meisten dieser Herausforderungen lässt sich bereits mit bestehenden Lösungen begegnen. Die Handhabung des Datenschutzes ist eine grundsätzliche Frage, wobei hier die Handhabung unternehmensintern und extern nicht variieren sollte und der Gesetzgeber hier bereits sehr klare Vorgaben macht. Bezüglich der Zusammenarbeit von Anbieter und Nutzer sind vertragliche Lösungen, wie SLA's, Nutzungsbedingen, Zertifizierungen etc. ein Instrument zur gegenseitigen Absicherung. So lassen sich Verlässlichkeit und eine verbindliche Haftung etablieren. Auch Transparenz über Datenherkunft und -Nutzung sowie die Implementierung von Standards und die gute Dokumentation auf beiden Seiten sind als integrale Instrumente des Sicherheits- und Compliance-Managements zu verstehen. Als Erfolgsbeispiel und Vorbild können Open-Source-SW-Lösungen dienen.

Wie kann Open Data bzw. Open API zu open Innovation führen und damit zum Wettbewerbsfaktor werden?

Open Data/Open API ist mit einer Vielzahl von Innovationsvorteilen verbunden. Zunächst entsteht über die strukturierte Öffnung große Transparenz bezüglich verfügbarer Services. Durch das stark vereinfachte Einbinden von Services können Parallelstrukturen abgebaut, Synergien genutzt und sogar gänzlich neue Geschäftsmodelle aufgebaut werden. Die Konzentration auf die Bereitstellung einzelner Services ermöglicht es, dass sich Unternehmen auf ihr Kerngeschäft konzentrieren und unnötige Kosten reduzieren. Die Öffnung für externe Nutzer eröffnet zudem die Möglichkeit für neue Nutzungskontexte, eine bessere Sichtbarkeit und Kundengewinnung sowie kontinuierliche Verbesserung durch Messen und Feedback. Zusammen mit der Nutzung gemeinsamer Ressourcen führt dies zu einer Win-Win-Situation für alle Beteiligten.

Damit die Vorteile von Open Data/Open API wirklich ausgeschöpft werden können und die Öffnung nicht zur überraschenden Disruption führt, müssen sich Unternehmen auf diese Transformation einstellen und sie aktiv vorantreiben. Ohne entsprechende Maßnahmen drohen Verlust von Kontrolle und bestehenden Wettbewerbsvorteilen durch exklusiven Zugriff auf die eigenen Daten und Services.

Letztendlich wird Open Data/Open API durch die einfache und unkomplizierte Verknüpfung von Services zu Innovation innerhalb wie außerhalb des Unternehmens führen, von der alle Beteiligten profitieren werden.

4. Weitere Informationen

Auch für das Jahr 2018 ist die Durchführung eines ESAPI-Workshops vorgesehen, der in München stattfinden wird. Weiterführende Informationen werden zeitnah unter der folgenden URL im Internet bereitgestellt:

https://blog.hwr-berlin.de/schmietendorf/

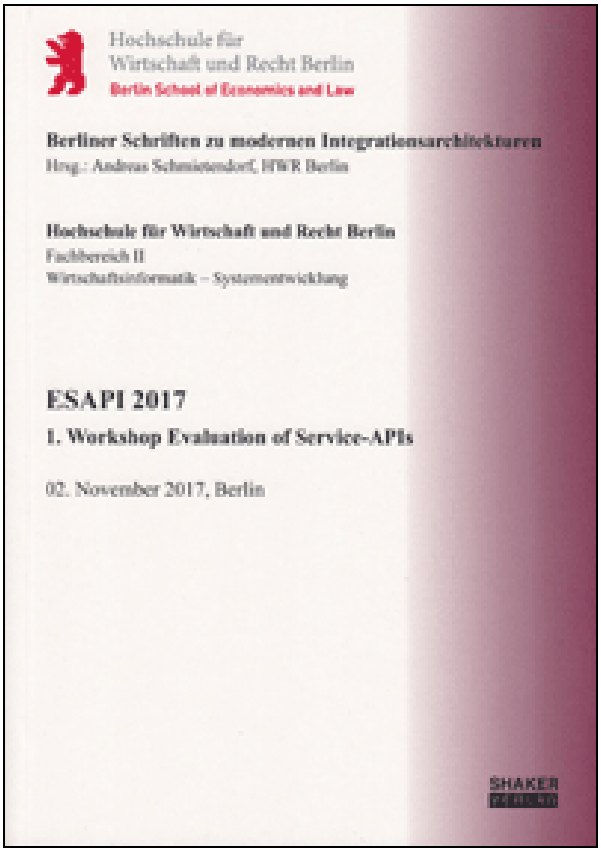


Abbildung 1: *Tagungsband zum Workshop ([Schmietendorf 2017])*

5. Quellenverzeichnis

[Schmietendorf 2017] Schmietendorf, A. (Hrsg.): ESAPI 2017 - 1. Workshop Evaluation of Service-APIs, in Berliner Schriften zu modernen Integrationsarchitekturen, Shaker-Verlag, Aachen, November 2017

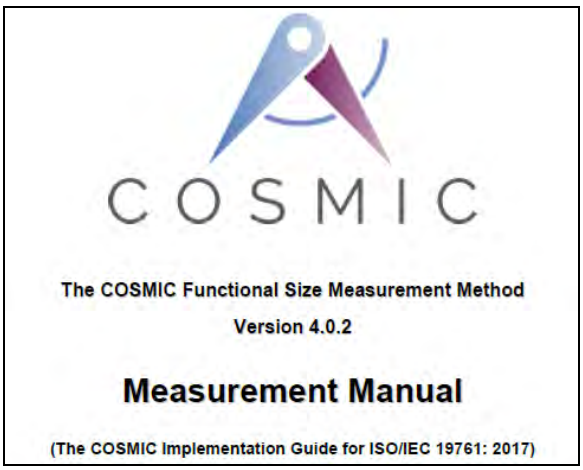
[Bitkom 2017] Leitfaden Open Data - Neue Konzepte erfolgreich umsetzen, URL: https://www.bitkom.org/Bitkom/Publikationen/Open-Data-Neue-Konzepte-erfolgreich-umsetzen.html, Nov. 2017

6 Dank

Unser Dank gilt den Referenten und Teilnehmern, aber auch den Sponsoren (Deutsche Bahn AG, T-Systems International GmbH und SEVEN PRINCIPLES AG), die eine solche Veranstaltung überhaupt erst ermöglicht haben. Ein herzlicher Dank geht auch an die beteiligten Medienpartner SIGS DATACOM GmbH aus Köln, dpunkt.verlag GmbH aus Heidelberg und an den Shaker Verlag GmbH aus Aachen.

COSMIC News

An extension of the COSMIC method 4.0 was established at October 2017 (see ***www.cosmic-sizing.org***):



Main changes for version 4.0.2 of the COSMIC Method

This v4.0.2 introduces no changes in the method's basic principles. The only changes affect definitions and rules and are all designed to improve ease of understanding and measurement repeatability. The most important changes occur in sections 3.2 and 3.3, and originate from Method Update Bulletins (MUBs) which have already been published, as follows.

The definition of 'Triggering event' has been modified to make clear that only the first data group generated by any one functional user will be moved by a triggering Entry. The possibility of a 'dual role' of a functional user (it may also be an object of interest) is also made clear in the definition of a functional user (MUB 12).

A new rule to 'identify different data groups (and hence different objects of interest) moved in the same one functional process' has been added (MUB 13)1.

In the definition of 'object of interest' the phrase 'process and/or move data' has been replaced by 'move a data group in or out of the software, or to or from persistent storage'. This change aims to make clear that an object of interest may be identified only if it is the subject of a data group that is moved. An object of interest may not be identified as a subject of only data manipulation. (MUB 14).
See Appendix E for a detailed list of all changes made from version 4.0 to v4.0.1 and from v4.0.1 to v4.0.2. This Measurement Manual for version 4.0.2 of the method becomes the current standard definition of the method from its date of publication.

Consequences of the main changes to v4.0.2 on existing size measurements, etc.

The original basic principles of the COSMIC method have remained unchanged since they were first published in the first draft of the Measurement Manual in 1999. This is in spite of the various refinements and additions needed to produce the International Standard and to produce all versions of the method up to this latest version 4.0.2.

Functional sizes measured according to the principles and rules of version 4.0.2 of the Measurement Manual may differ from sizes measured using earlier versions only because the new rules intend to be more precise and complete. Hence Measurers have less discretion for personal interpretation of the rules than was possible with earlier versions.

As a further indication of the continuity of the method, anyone who passed the Foundation level certification examination for version 3.0/3.0.1/4.0 of the method will be considered to be still certified for v4.0.1 and v4.0.2 of the method at the Foundation level.

Impact Analysis of Multiple Imputation on Effort Estimation Models with the ISBSG Repository of Software Projects

Abdalla Bala[a], Alain Abran[b]

[a]abdalla.bala@techbala.com, Bala Group Technologies Inc., Montréal, Québec, Canada.

[b]alain.abran@etsmtl.ca, Ecole de Tehcnologie Superieur (ETS)
Univerity of Quebec, Montreal, Quebec, Canada

Abstract

BACKGROUND: The ISBSG data repository contains software project data collected from various organizations around the world.

PROBLEM: Software engineering data sets, such as the ISBSG repository, typically contain a large number of missing data, which considerably reduces the number of data points available for building estimation models. What is the impact on the accuracy of estimation models built from imputed datasets when simple and multiple imputation techniques are used to replace missing data?

METHOD: This study used both simple and multiple imputation techniques with either absolute or relative seeds, to build effort estimation models and then applied regressions on these imputed datasets to compare their respective performance. The performance of imputation techniques was investigated using data from projects within the ISBSG dataset release 9with data for all project phases divided into two subsets: subset X consisting of complete data fields for all project phases and subset Y consisting of information from which implementation activity data is deleted. Subset X was used as the training dataset for the imputation and for building the estimation model from multiple regression analysis, which was then applied to subset Y with the missing values. The performance of the effort prediction models, both estimated effort and actual effort, were then evaluated using three evaluation criteria: adjusted R^2, MMRE, and Pred(25).

RESULTS: The key findings are:

1. When only the absolute value was calculated, multiple imputation with absolute seeds performed much better in comparison to simple imputation with an absolute average.
2. When a value of the variable being estimated was calculated relative to other fields, then simple imputation with a relative average seed performed best for all techniques and types of seeds tested.
3. When estimates were compared to the complete dataset:
   - Simple imputation with a relative average seed performed best for all techniques and types of seeds tested, with estimation results almost as good as estimates from the complete dataset on all three criteria of adjusted R^2, MMRE, and Pred(25).
   - MI with absolute seeds was the next best strategy in terms of adjusted R^2 and Pred(25), but degraded MMRE.

CONCLUSION: Our research methodology and results allow software researchers and practitioners to assess and compare the accuracy of estimation models built from imputed data, and provides practical guidelines for using imputation techniques on data sets with a large number of missing values.

Keywords: *missing data; imputation techniques; ISBSG; software effort estimation; estimation model; evaluation criteria.*

1. Introduction

Multi-organizational repositories, in particular those based on voluntary data contributions, such as the repository of the International Software Benchmarking Standards Group (ISBSG), may be missing a large number of values for many of their data fields, as well as including some outliers.

(Bala & Abran 2016) identified a number of data quality issues associated with the ISBSG repository that can compromise outcomes for users exploiting it for benchmarking purposes or for building estimation **models.** With conventional statistical methods, all the variables in a specified model are presumed to be collected and available for all cases. The default action for virtually all statistical tools is simply to delete cases with any missing data on the variables of interest, a method known as *listwise deletion* or *complete case analysis*. While this simple treatment might be acceptable with a large dataset and a relatively small amount of missing data, biased findings can result if the percentage of missing data is significant, as information on the incomplete cases will have been lost. With relatively small datasets, it is poor practice to ignore missing values or to delete incomplete observations in these situations. The most obvious drawback in listwise deletion is that it often removes a large fraction of the sample, which results in a serious loss of statistical power. More reliable imputation methods must be used to ensure that the analyses in which they are used are meaningful. Awareness of the importance of treating missing data in appropriate ways during analysis has been growing (Idri, Abnane & Abran 2016).Consequently, techniques for dealing with missing multivariate data have been proposed, including the use of the multiple imputation (MI) technique, where instead of filling in a single value for each one missing, an MI procedure replaces each missing value with a set of plausible values that represent the uncertainty about the right value to impute(Rubin 1987)(Little 1998) (Graham & Schafer 1999) (Little & Rubin 2002). These multiple imputed data sets can then be analyzed using standard procedures for complete data and combining the results from these analyses.

(Bala & Abran 2016) investigated the use of MI to deal with missing values in the ISBSG repository including the implication of the presence of outliers in numerical data fields (Abran 2015), but did not investigate the impact of MI when building effort estimation models from imputed datasets. This paper reports on a follow-up study to specifically investigate the performance of effort estimation models built from the ISBSG dataset where missing effort values have been imputed using the MI technique. The analysis presented is relevant to software engineering repositories in general and is not limited to the ISBSG repository.

The paper is organized as follows. Section 2 presents related work. Section 3 presents the research methodology, including the data preparation on ISBSG R9 for effort by project phase. Sections 4 and 5 present the outcomes of two options, absolute seeds and relative seeds, for verification of the contribution of the MI technique on effort estimation. Section 6 presents a simple imputation and MI on a set of 40 projects from a previous study with both absolute and relative seeds, together with effort estimation from the imputed datasets. Section 7 compares effort estimate performance for all distinct imputations techniques and strategies. Section 8 presents the conclusions and suggested future work.

2. Related Work

(Mockus2008) (Catwright *et al.* 2003) (Deng & MacDonell 2008) (Song & Shepperd 2007) applied missing data techniques on a software engineering data set to illustrate a variety of practical contexts where such techniques were needed and to highlight the pitfalls of ignoring the missing data problem. The most common approaches to handling missing data, along with a brief description, as well as the drawback of each approach are:

1. Listwise deletion: Deletes an entire observation if it is missing on any case used in the analyses.

2. Mean Imputation: Replaces a missing observation with the sample mean of the variable. To reduce the influence of exceptional data, the median is used. Using such a constant to replace missing data will change the characteristics of the original dataset and ignoring the relationship among attributes will bias the data. A variation of this method is to replace the missing data for a given attribute by the mean or mode of all known values of that attribute in the class where the instance with missing data belongs (Magnani 2004).
3. Pairwise Deletion: Uses all available data between pairs of variables to calculate covariance by only deleting "pairs" with specific missing data (e.g., not the whole case is deleted). This only works if a model is estimated from covariance or correlation matrices.
4. Hotdecking: A missing attribute value is filled in with a value from an estimated distribution for the missing value from the current data (Acuna & Rodriguez 2004). It is typically implemented in two stages: data are partitioned into clusters, and then missing data are replaced within a cluster by calculating the mean or mode of the attribute within the cluster.
5. Single imputation: Substituting model-predicted values instead of missing data. Typically based on a multivariate model using the same variables as in the main analysis.
6. Multiple Imputation (MI): assumes that missing data are missing at random, that is, missing values do not carry any extra information about why they are missing than what is already available in the observed data. MI does not attempt to estimate each missing value through simulated values but rather represents a random sample of the missing values. The objective of MI is not to predict missing values that are as close as possible to the true values, but to handle missing data in a way that results in valid statistical inference. This process results in valid statistical inferences that properly reflect the uncertainty due to missing values, e.g. has valid confidence intervals for parameters. MI inference involves three distinct phases (Rubin 1987):
    - The missing data are filled in *m* times to generate *m* complete datasets.
    - The *m* complete datasets are analyzed using standard procedures.
    - The results from the *m* complete datasets are combined for the inference.

(Strike *et al.* 2001)carried out a simulation to evaluate three techniques for dealing with missing data in software cost modeling: listwise deletion, mean imputation, and eight different types of hot-deck imputation. (Myrtveit *et al.* 2001)evaluated four missing data techniques for software cost modeling: listwise deletion, mean imputation, similar response pattern imputation, and full information maximum likelihood. They applied these techniques to an ERP data set, and then constructed regression-based prediction models using the resulting data sets.(Horton & Lipsitz 2001)evaluated a number of software packages that implement an MI procedure including interfaces, features, and shortcomings.

The research goal for (Bala & Abran 2016) was to develop improved usage of the ISBSG data repository by both practitioners and researchers by leveraging the larger quantity of data available for statistical analysis in software engineering, while discarding the data that could affect the meaningfulness of the statistical tests. To achieve this research goal of tackling the new problems in larger datasets in software engineering including outliers and missing values, they investigated the impact of MI on the estimation of missing values of the effort variable by project phase using the ISBSG repository, and applied regression models, both with and without outliers, and examined their specific influence on the results. (Bala & Abran 2016) used Release 9 (R9) of the ISBSG data repository with a statistical model to explain the variability in the 'total project effort'. They also investigated the impact of MI with five imputation rounds in the estimation of the missing values of the 'effort by project phase' using the 179 projects from the ISBSG R9 repository with the necessary detailed effort information by project phases as defined by ISBSG:

- P: Plan
- S: Specify
- B: Build (e.g. coding)
- T: Test
- I: Implement.

Two sets of data points had enough detailed effort data for statistical analysis:

- PSBTI profile with no missing value for each of the five phases of P, S, B, T and I.
- PSBT profile has data only for the P, S, B and T phases (e.g. has missing data for the I phase).

Findings from the MI study of (Bala & Abran 2016) were not evaluated and compared with previous studies using a simpler imputation technique, such as mean imputation.

3. Research Methodology

Since the purpose of this study is to compare the results of MI with the simpler imputation method of (Déry & Abran 2005), the dataset selected consists of the same 41 projects from ISBSG R9 with complete data values for the PBSTI phase profile and an imputation of the Effort in Implementation (EI) phase and its subsequent estimation as the dependent variable in regression models.

The research methodology selected for analyzing the performance of MI consists of artificially creating a subset by deleting a number of data values, and then comparing the estimation models of EI derived from the original dataset and from the MI applied to the artificial subset with EI missing data. The specific strategy adopted in this study consists of:

- Randomly splitting the (Déry & Abran 2005) data set into two subsets X and Y.
- deleting the data values for the EI data field from subset Y,
- Replacing these deleted data with imputed values in subset Y.
- Building EI estimation models with both the initial complete data set and the imputed dataset.
- Assessing the predictability of these EI estimation models based on the following criteria (Conte *et al*. 1986):
  - Magnitude of Relative Error (MRE) = | Estimated value – Actual value | / Actual
  - Mean Magnitude of Relative Error for *n* projects (MMRE) = 1/n*Σ(MREi)
  - Measure of Prediction Quality = Pred(x/100) – here Pred(25).

The new study reported here looks at two options for imputing data based on MI by random selection from min-max seeds:

- Option A: Using MI with seeds from the absolute min-max values of EI is referred to as imputation from absolute EI values. Results of option Aare presented in Section 4.
- Option B: Using MI with seeds from relative min-max values of EI values (e.g., when the seeds are the min-max values of EI relative to Total Effort. Results of option B are presented in Section 5.

Next, the same dataset with same sample of projects with a PSBTI profile from the (Déry & Abran 2005) study with imputed data from either the simple average or relative values of the EI on the results are presented in Section 6.

Comparison of the effort estimates of all of the above are presented in Section 7.

4. Option A: Random imputed values from absolute EI seeds

4.1 Strategy for Option A (N=41)

The strategy for option A for investigating the contribution of the MI technique, given an initial dataset without missing values (see Figure 1)- consists of:

a) Randomly selecting projects and splitting the initial complete data set into two subsets (X and Y).
b) Creating missing values artificially in subset Y by randomly deleting data from the EI phase data field. Selecting seeds from the absolute EI values for min & max from subset X.

c) Assigning random values to subset Y and creating five imputed datasets (e.g., by combining subsets X and Y imputed).
d) Combining imputation results.
e) Building a regression model to estimate EI based on the other four project phases for:
   - Estimate of EI with the initial complete dataset.
   - Estimates of EI with the dataset with EI imputed values.
f) Comparing estimates by assessing and comparing the predictability with MMRE and Pred(25).

4.2 Impact of MI on parameter estimates (N=41)

This section presents the results of the estimation model for the dependent variable EI trained with the independent variables (Effort Plan, Effort Specify, Effort Build, and Effort Test) – see Table 1:

- For the complete dataset of 41 projects with data in the PSBTI profile (first line, Table 1)
- For each of the five imputations (for the 41 projects, including the MI for the 21 missing values)
- For the combined imputation model (bottom line, Table 1).

Since this estimation model was based on a multi-regression with four (4) independent variables (EP, ES, EB and ET), it was preferable to use the adjusted R^2 to the R^2 for evaluation purposes. Here, the adjusted R^2 for the estimation models derived from the combined imputations was 0.45 (bottom line of Table 1), as compared to the adjusted R^2 of 0.58 for the estimation model from the complete dataset (top line of Table 6.1).

Table 1 presents the P-value for the complete dataset (<0.0001) for each of the five imputations as well as the P-value for the combined imputations (<0.0001). The P-values are all less than the 0.1 criterion for a P-value indicating they are statistically significant (i.e. a P-value of 0.05 means that there is a 5% chance that the relationship is real at the 95% confidence level for a P-value of less than 0.05).

4.3 Impact on parameter estimates excluding outlier (N = 40)

The outlier identified using the Grubb test of (Bala & Abran 2016) was excluded from the data set (e.g., project id (9): a size = 2189 function points and an EI = 117 hours). Table 2 presents the results of the estimation model for the dependent variable EI trained with the independent variables Effort Plan, Effort Specify, Effort Build, and Effort Test for each of the five imputations, based on N=40 projects (i.e. excluding an outlier).

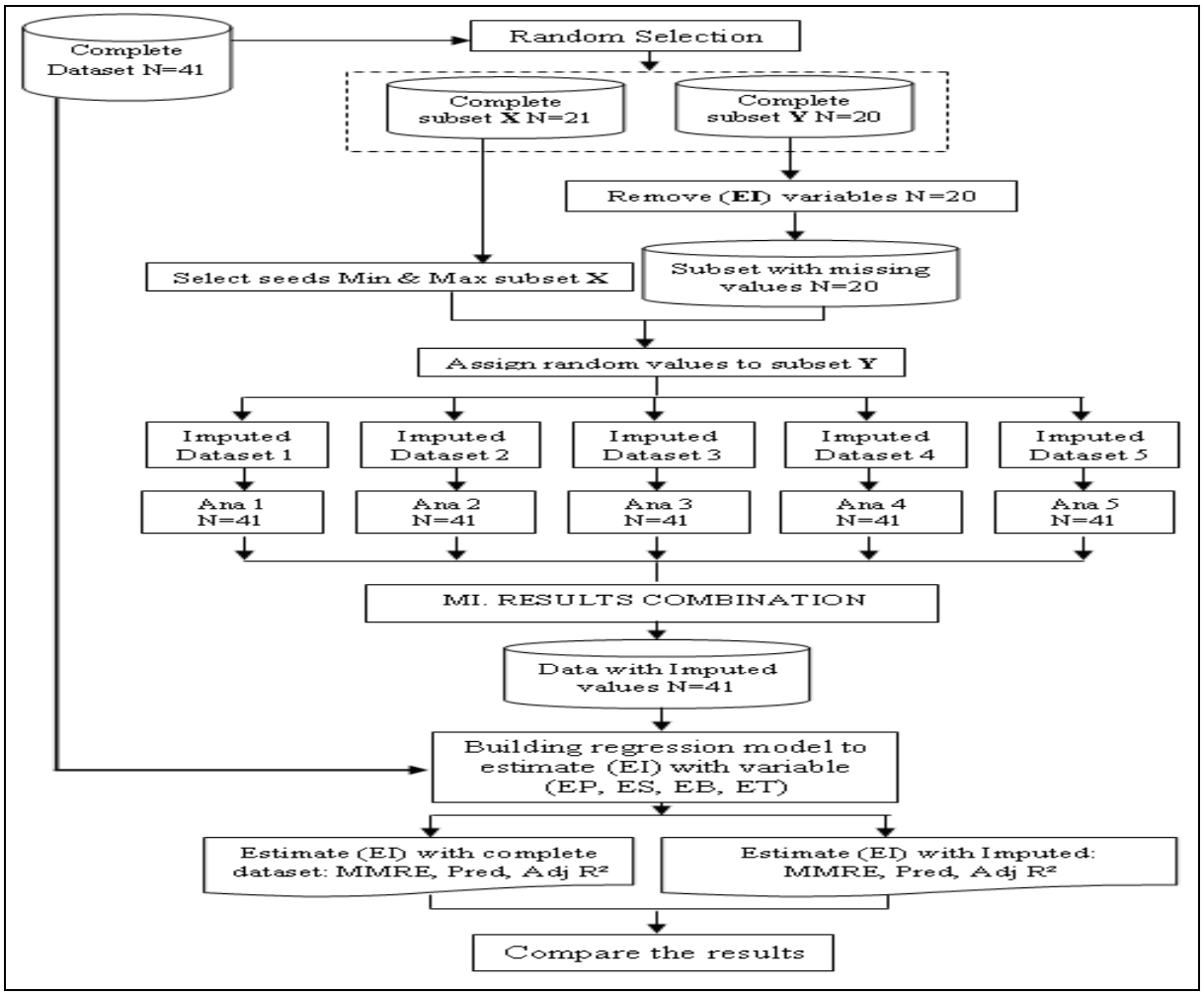


Figure 1: Option A: absolute seeds - analyzing the predictive accuracy of an MI dataset (N = 41)

Table 1: MI with absolute EI seeds – estimation models for EI (N=41 projects)

Dataset	Intercept	Effort Plan	Effort Specify	Effort Build	Effort Test	Adjusted R^2	R^2	P-value
Complete dataset	45	0.57	0.10	-0.001	-0.06	0.58	0.62	<0.0001
MI 1	368	0.49	-0.04	0.06	-0.05	0.36	0.42	0.0004
MI 2	246	0.49	0.07	0.009	-0.09	0.39	0.45	0.0002
MI 3	278	0.60	0.12	0.02	-0.22	0.46	0.51	<0.0001
MI 4	227	0.52	0.08	0.007	-0.10	0.41	0.47	<0.0001
MI 5	154	0.53	0.02	0.02	0.02	0.55	0.59	<0.0001
Combined imputations	255	0.53	0.05	0.02	-0.09	0.45	0.46	<0.0001

Table 2: MI with absolute EI seeds – estimation models for EI (N=40)

Dataset	Intercept	Effort Plan	Effort Specify	Effort Build	Effort Test	Adjusted R^2	R^2	P-value
Complete dataset	-7	0.67	0.15	-0.06	0.03	0.71	0.74	<0.0001
MI 1	169	0.61	0.08	-0.04	0.05	0.52	0.57	<0.0001
MI 2	46	0.68	0.07	-0.04	0.13	0.71	0.74	<0.0001
MI 3	203	0.65	0.15	-0.04	-0.09	0.55	0.59	<0.0001
MI 4	41	0.66	0.11	-0.07	0.12	0.71	0.74	<0.0001
MI 5	-6	0.71	0.06	-0.06	0.23	0.76	0.78	<0.0001
Combined imputations	91	0.66	0.09	-0.05	0.09	0.65	0.69	<0.0001

The estimation model based on the complete data set of 40 projects is:

EI = -7hrs + 0,67xEP + 0,15xES – 0,06xEB + 0,03xET (line 1, Table 2)

The estimation model based on the combined estimation is:

EI = -91hrs + 0,66xEP + 0,09xES – 0,05xEB + 0,09xET (bottom line, Table 2)

Table 2 presents the adjusted R²forthe complete dataset (adjusted R² = 0.71) for each of the five imputations (0.52, 0.71, 0.55, 0.71, 0.76), and the combined imputation (0.65).This means that, after removing a single outlier, the adjusted R² increased in each of the five imputations, as well as in the combined imputation, and comes even closer to the adjusted R² for the complete dataset (i.e. 0.71). All the models hada significant P-value <0.0001indicating they are statistically significant.

In summary, removing the outlier strengthened the linearity of the data and decreased the errors present in the regression. Furthermore, the results were statistically significant for the estimates of EI, as illustrated by the t-test and P-values with and without outliers.

Table 3 presents an analysis of the variance between estimated EI and actual EI using MMRE and Pred(25) to assess the results of the regression models for the five imputed datasets and combined imputations. These statistics made it possible to compare the performance of the estimation model based on combined imputations of MI for half of the dataset with the performance of the estimation model based on the complete dataset(with and without outliers).

Table 3: MI with absolute EI seeds – comparison of EI estimation models (with & without an outlier)

Dataset	N=41 projects		N=40 projects -excluding an outlier	
	MMRE	Pred(25)	MMRE	Pred(25)
Complete dataset	110%	32%	88%	30%
MI 1	259%	24%	148%	18%
MI 2	173%	22%	101%	28%
MI 3	192%	15%	149%	23%
MI 4	169%	24%	82%	33%
MI 5	143%	22%	101%	33%
Combined imputations	187%	23%	116%	27%

Even though the adjusted R² was relatively high (0.71 in Table 2) for the EI estimation model for this set of 40 projects without an outlier, Table 3 shows that, for the complete dataset, the quality of the estimation model built was not particularly high, with MMRE = 88% and Pred(25) = 30%(line 1 in Table 3).

With 50% of the data missing (i.e. 20 missing values in a sample of 40 projects excluding one outlier), a much larger MMRE error and worse Pred(25) would be expected. However, with MI the regression results of the combined imputation model showed only a relatively minor reduction in quality, where the MMRE at 116% was not that far from that of the complete dataset with MMRE at 88 %,and the Pred(25) at 27% very close (within 3%) (bottom line in Table 3).

4.4 Comparison of EI effort estimation (from the training dataset without an oulierN=20)

This section compares the MI results with the results of the EI estimation model derived from the training dataset X of 20 projects (instead of the estimation model derived from the complete dataset of 40 projects excluding the outlier)and divides it into two subsets – see Figure2.

- Subset X of 20 projects with complete data fields;
- Subset Y of the other 20 projects from which the information in the EI data field is deleted (this will be referred to as Subset Y with missing EI data).

This section builds estimation models with only subset X with the PSBTI profile used as the training dataset, and then applies this estimation model to subset Y. Figure 2 illustrates this specific strategy, given the PSBTI dataset (N= 40 projects excluding one outlier):

a) Build an estimation model for EI with subset X – the training data set

b) Apply the EI estimation model from a) on subset Y.

c) Analyzethe EI estimation variance on the subset Y with imputed training dataset to assess the predictability with MMRE and Pred(25) on subset Y.

d) Build an estimation model from the five combined imputation datasets of MI N= 20 projects with missing values.

e) Analyzethe EI estimation variance with the five combined imputation datasets of MI N=20 projects to assess the predictability with MMRE and Pred(25).

f) Compare the results of MMRE and Pred(25) of subset Y with the results from the training dataset (e.g., subset X), and the combined imputation datasets of MI.

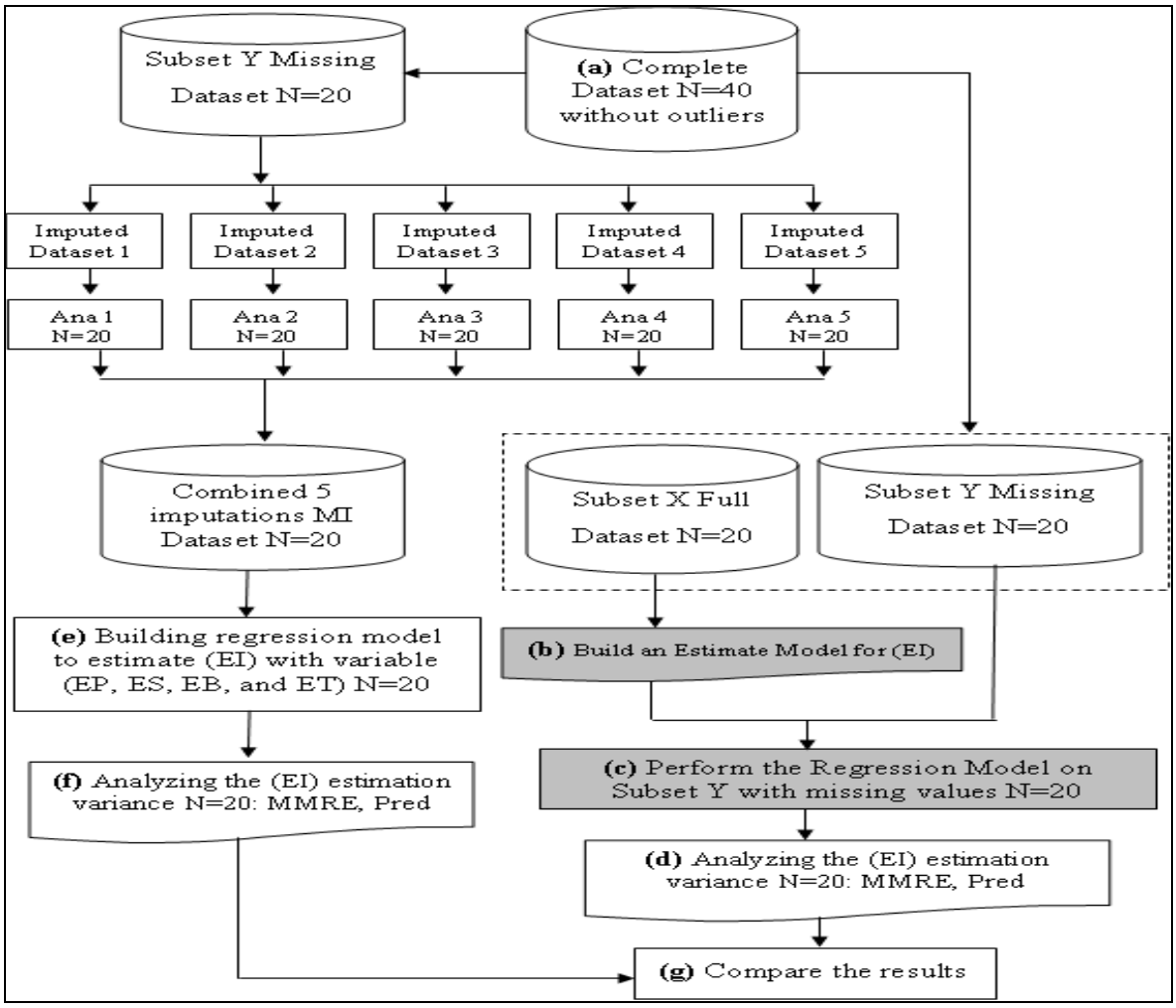


Figure 2: Option A – MI with absolute EI seeds - analyzing the predictive accuracy of an MI dataset (N = 40)

Table 4 presents the results of two multi regression EI estimation models built from:

- The training subset X.
- The combined imputed data set (40 projects from subset X and the imputed data on subset Y).

The EI estimation model based only on the training subset A of 20 projects is:

EI = -59hrs + 0,78xEP + 0,16xES – 0,1xEB + 0,11xET(line 1 in Table 4).

Table 4: MI with absolute EI seeds – estimation models for EI (N=20, excluding an outlier)

Dataset	Intercept	Effort Plan	Effort Specify	Effort Build	Effort Test	Adjusted R^2	R^2	P-value
Training subset X (N=20)	-59	0.78	0.16	-0.1	0.11	0.69	0.76	0.0002
Combined imputations MI	247	0.23	0.07	-0.03	0.20	0.37	0.39	<0.0001

Table 4 shows that for the estimation models built:

- For the training subset X the adjusted R^2 = 0.69.

- For the combined 5 imputations of MI the adjusted R^2 = 0.37.

Table 5: MI with absolute seeds - EI estimation variance from estimation with imputed variance and training estimation model

No.	No. projects, without outliers	MMRE EI	Pred(25) EI
1	Training subset X N=20	97%	30%
2	Combined imputations MI N=20	92%	40%
3	Subset Y with missing values N=20	118%	15%
	Comparison results vs. complete data		
4	**2 vs. 1**	-5%	+10
5	**3 vs. 1**	+21%	-15
6	**2 vs. 3**	-26%	+25

Table 5 shows that:

- For the estimation model derived from the training subset X (N=20) MMRE = 97% and Pred(25) = 30%– line 1 in Table 5.
- For the estimation model derived from the five combined imputations of MI MMRE = 92% and Pred(25) = 40%– line 2 in Table 5.
- MMRE =118% and Pred(25) =15% for the performance on subset Y derived from the estimation model produced from training subset X – see line 3 in Table 5.

Compared to the performance of the estimation model built with training subset X the performance of the combined imputation results of MI represents a 5%decrease in MMRE, and a 10%increase in Pred(25), which is an improvement on both performance criteria – line 4 in Table 5.

5. Option B: random imputed values from relative EI seeds

There is often a considerable variation of effort by phase at the project level and the ratio of EI to Total Effort may vary considerably across projects. For example, in this dataset of N=40 projects (excluding an outlier), the minimum EI percentage was 1% of Total Effort and the maximum EI percentage was 41% of Total Effort. However, this absolute max of 2,946 EI hours represents only 24% of the Total Effort for this project, far lower than the maximum of 41% of relative EI for the whole set of 40 projects. This section examines the impact of changing the EI seed values from the absolute min and max values of EI to their relative in and max of EI with respect to 'Total Effort'. Altering the basis for the imputation by changing the seed values is, in effect, a sensitivity analysis.

To accomplish this, the set of 40 projects with the PSBTI profile was used again and the relative value of EI for the project with this absolute min and max determined as follows:

a) Identify the project with the absolute Min and Max EI.

b) Calculate for this specific project the maximum percentage of EI to Total Effort for theproject using the following equation:

$$\text{Relative } (\%EI) = (\text{absolute } EI_i / \sum_i (EP_i + ES_i + EB_i + ET_i + EI_i)) \times 100$$

Figure 3 illustrates our specific strategy for investigating the sensitivity of the analysis with the results of MI technique, given the dataset of the PSBTI profile without missing values N=40 projects:

a) Create missing values artificially by deleting data from the EI data field in subset B.

b) Identify and calculate the Relative EI $(\%EI) = (\text{absolute } EI_i / \sum_i (EP_i + ES_i + EB_i + ET_i + EI_i)) \times 100$.

c) Generate random numbers based on the relative seeds from N=40 projects.

d) Select the seeds (here: Min= 1%, Max= 41%) from relative min-max values of (EI).

e) Build an estimation model from the relative imputation results.

f) Select the seeds (here: Min= 1%, Max= 24%) from absolute min-max values of (EI).

g) Build an estimation model from the full dataset based on the seeds selected in f).

h) Compare relative imputation results of MMRE and Pred(25) with the original dataset without missing values.

Table 6 presents the multi-regression models as well as their adjusted R^2 (N=40) for the:

- Complete dataset: 0.71
- Combined imputation with relative seeds (%Min = 1%, %Max = 24%: 0.60
- Combined imputation seeds with relative seeds (%Min = 1%, %Max = 41%): 0.53
- Combined imputation from absolute seeds: 0.65 – as in Table 2.

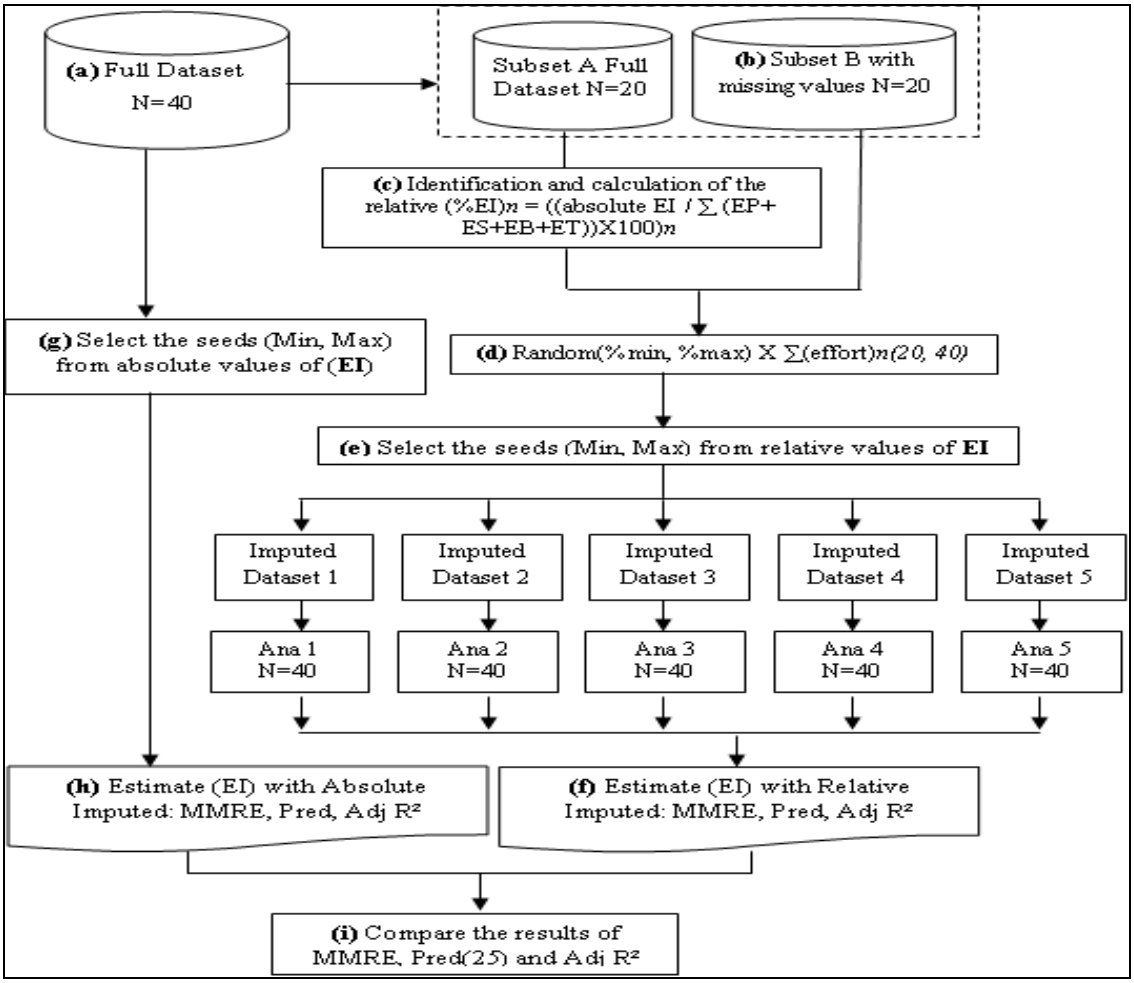


Figure 3: Option B: Relative EI seeds - Analyzing the predictive accuracy of an MI dataset

Table 6:MI with relative EI seeds – Estimation models for EI (N=40)

Dataset	Intercept	Effort Plan	Effort Specify	Effort Build	Effort Test	Adjusted R^2	R^2	P-value
Complete dataset	-7	0.67	0.15	-0.06	0.03	0.71	0.74	<0.0001
Combined imputed (relative seeds %1 to %24)	82	0.59	0.08	-0.01	0.06	0.60	0.64	<0.0001
Combined imputed (relative seeds %1 to %41)	174	0.54	0.003	0.01	0.13	0.53	0.58	<0.0001
Combined imputations of MI absolute seeds	91	0.66	0.09	-0.05	0.09	0.65	0.69	<0.0001

Table 7 presents the quality of the estimation models for:

- The complete dataset: MMRE = 88% and Pred(25) = 30%
- With relative imputation for seeds (1% to 24%):MMRE = (147%) and Pred(25) = 24%
- The relative imputation for seeds (1% to 41%):MMRE = (172%) and Pred(25) = 23%
- The combined results of MI from absolute seeds: MMRE = 116% and Pred(25) = 27% - as in Table 3.

Table 7: MI with relative EI seeds - Contribution of relative imputation for N=40

Line No.	Imputation no.	%MMRE	Pred(25)
1	Complete dataset	**88%**	**30%**
2	Combined imputed (relative seeds %1 to %24)	**147%**	**24%**
3	Combined imputed (relative seeds %1 to %41)	**172%**	**23%**
4	Combined imputations of MI absolute seeds	**116%**	**27%**
Comparison results vs. complete dataset			
6	2 vs. 1	**+59%**	**-6%**
7	3 vs. 1	**+84%**	**-7%**
8	4 vs. 1	**+28%**	**-3%**

In summary, with the relative imputation, Table 7shows that in comparison to the MMRE and Pred(25) of the EI estimation from the complete dataset, the difference for the combined models imputed with:

- Relative seeds (%Min = 1% and %Max = 24%) increases MMRE by + 59%, and decreases Pred(25) by -6% – line 6 in Table 7.
- Relative seeds (%Min = 1% and %Max = 41%) increases MMRE by 84%, and decreases Pred(25) by -7% – line 7 in Table 7.
- Absolute seeds increases MMRE by 28%, and decreases Pred(25) by -3% – line8 in Table 7.

6. Imputation and EI estimation from simple imputation and MI

This section presents simple imputations and MI based on absolute and relative averages using the same dataset as (**Déry & Abran 2005)**.

6.1 Simple imputations from absolute and relative averages

(Déry & Abran 2005) imputed the missing values of EI in the PSBT profile with missing EI using the average% of EI from the PSBTI profile (a constant). This approach does not permit verification of the performance of estimation of EI for the PSBT profile with respect to their actual values since the values for EI were missing in this PSBT profile. Therefore, to analyze the estimation performance of average imputation for missing values, the same strategy adopted in the previous sections was used here.

As with the previous strategies, this section selects the same 40 projects (excluding one outlier) with the PSBTI profile with its complete dataset, and division into the same two subsets:

- Subset X of 20 projects with a PSBTIprofile , with complete data values, and
- Subset Y of 20 projects with a PSBT profile with missing EI obtained by deleting data from the data field of EI.

Table 8 presents the complete dataset N= 40 projects without missing values.

Table 8: Sample of complete data N=40 projects

No	EP	ES	EB	ET	EI
1	3120	1540	5260	2299	2946
2	208	624	2075	761	528
3	2	38	91	15	24
4	80	80	80	40	20
5	1190	9793	17167	4489	1384
6	578	3412	3768	1540	1320
7	78	1435	640	1794	374
8	196	565	204	163	466
9	527	2278	1016	2335	260
10	94	1240	858	905	234
11	78	546	312	312	156
12	30	694	330	550	100
13	468	3160	1423	1493	577
14	283	4444	2837	1419	473
15	83	2339	826	880	606
16	78	437	156	109	78
17	47	1624	757	519	39
18	750	120	2110	800	140
19	2293	1048	7193	891	563
20	100	140	840	380	40
21	1810	156	4016	3014	790
22	150	310	1402	125	200
23	322	161	1979	622	168
24	140	242	414	140	23
25	120	288	592	372	92
26	330	200	380	400	90
27	140	100	4083	1025	82
28	85	255	1575	708	310
29	554	850	1215	3290	543
30	806	875	989	726	243
31	755	2976	5860	331	654
32	16	47	31	93	47
33	160	160	400	303	61
34	234	468	468	312	78
35	213	221	1864	296	150
36	16	156	311	117	47
37	300	520	1300	390	120
38	312	1248	1014	780	390
39	48	82	512	52	20
40	31	733	312	734	156

Table 9 displays the same dataset with N=40 projects with missing values for subset Y, that is:

- Subset X consisting of20 complete projects (PSBTI profile);
- Subset Y consisting of the 20 projects with missing effort in the 'implement' phase (PSBT profile - see the shaded areas in Table 9).

For the set of 20 projects in subset X:

- Absolute average EI = 516 hours
- Absolute EI min = 20 hours
- Absolute EI max = 2946 hours
- Relative EI min = 1%
- Relative EI max = 41%

Table 9: Division of the 40 projects with data deleted in subset Y

Input	EP	ES	EB	ET	EI
1	3120	1540	5260	2299	2946
2	208	624	2075	761	528
3	2	38	91	15	24
4	80	80	80	40	20
5	1190	9793	17167	4489	1384
6	578	3412	3768	1540	1320
7	78	1435	640	1794	374
8	196	565	204	163	466
9	527	2278	1016	2335	260
10	94	1240	858	905	234
11	78	546	312	312	156
12	30	694	330	550	100
13	468	3160	1423	1493	577
14	283	4444	2837	1419	473
15	83	2339	826	880	606
16	78	437	156	109	78
17	47	1624	757	519	39
18	750	120	2110	800	140
19	2293	1048	7193	891	563
20	100	140	840	380	40
21	1810	156	4016	3014	
22	150	310	1402	125	
23	322	161	1979	622	
24	140	242	414	140	
25	120	288	592	372	
26	330	200	380	400	
27	140	100	4083	1025	
28	85	255	1575	708	
29	554	850	1215	3290	
30	806	875	989	726	
31	755	2976	5860	331	
32	16	47	31	93	
33	160	160	400	303	
34	234	468	468	312	
35	213	221	1864	296	
36	16	156	311	117	
37	300	520	1300	390	
38	312	1248	1014	780	
39	48	82	512	52	
40	31	733	312	734	

Table 10 displays the average effort distribution of work effort across development phases for the subsets X and Y and for the full dataset. The average effort distribution for the EI wascalculated based on:

EI% = (Total EI / Total Effort (P+S+B+T+I)) x 100EI% = 8.2% for subset X

EI% = 6.7% for subset Y

Table 10: Average effort distribution by project phase (N=40)

Imputation #no.	Profile	Project Phases – % Effort					No. of projects
		Effort Plan	Effort Specify	Effort Build	Effort Test	Effort Implement	
Complete data	**Subset X**	8.2	28.3	38.1	17.2	8.2	**20**
	Subset Y	10.3	15.8	45.3	21.8	6.7	**20**
	Subsets X&Y	8.9	24.1	40.5	18.8	7.7	**40**

For the imputation based on the absolute average of the 20 projects having missing values, the imputation was made only to the EI data points having missing values. For example, Table 11 presents the imputation with the absolute average of 516 hours for EI (e.g., the imputation of a constant value).

Table 11: Imputation to subset Y based on absolute average EI of subset X

Input	EP	ES	EB	ET	EI
1	3120	1540	5260	2299	2946
2	208	624	2075	761	528
3	2	38	91	15	24
4	80	80	80	40	20
5	1190	9793	17167	4489	1384
6	578	3412	3768	1540	1320
7	78	1435	640	1794	374
8	196	565	204	163	466
9	527	2278	1016	2335	260
10	94	1240	858	905	234
11	78	546	312	312	156
12	30	694	330	550	100
13	468	3160	1423	1493	577
14	283	4444	2837	1419	473
15	83	2339	826	880	606
16	78	437	156	109	78
17	47	1624	757	519	39
18	750	120	2110	800	140
19	2293	1048	7193	891	563
20	100	140	840	380	40
21	1810	156	4016	3014	516
22	150	310	1402	125	516
23	322	161	1979	622	516
24	140	242	414	140	516
25	120	288	592	372	516
26	330	200	380	400	516
27	140	100	4083	1025	516
28	85	255	1575	708	516
29	554	850	1215	3290	516
30	806	875	989	726	516
31	755	2976	5860	331	516
32	16	47	31	93	516
33	160	160	400	303	516
34	234	468	468	312	516
35	213	221	1864	296	516
36	16	156	311	117	516
37	300	520	1300	390	516
38	312	1248	1014	780	516
39	48	82	512	52	516
40	31	733	312	734	516

The shaded area in Table 12 presents the imputation results for the subset Y based on the relative % average of EI for each project. The imputed EI hours vary from a min of 15 hours to a max of 738 hours.

Table 12: Imputation to subset Y based on relative %average

Input	EP	ES	EB	ET	EI
1	3120	1540	5260	2299	2946
2	208	624	2075	761	528
3	2	38	91	15	24
4	80	80	80	40	20
5	1190	9793	17167	4489	1384
6	578	3412	3768	1540	1320
7	78	1435	640	1794	374
8	196	565	204	163	466
9	527	2278	1016	2335	260
10	94	1240	858	905	234
11	78	546	312	312	156
12	30	694	330	550	100
13	468	3160	1423	1493	577
14	283	4444	2837	1419	473
15	83	2339	826	880	606
16	78	437	156	109	78
17	47	1624	757	519	39
18	750	120	2110	800	140
19	2293	1048	7193	891	563
20	100	140	840	380	40
21	1810	156	4016	3014	738
22	150	310	1402	125	163
23	322	161	1979	622	253
24	140	242	414	140	77
25	120	288	592	372	113
26	330	200	380	400	107
27	140	100	4083	1025	439
28	85	255	1575	708	215
29	554	850	1215	3290	485
30	806	875	989	726	278
31	755	2976	5860	331	814
32	16	47	31	93	15
33	160	160	400	303	84
34	234	468	468	312	122
35	213	221	1864	296	213
36	16	156	311	117	49
37	300	520	1300	390	206
38	312	1248	1014	780	275
39	48	82	512	52	57
40	31	733	312	734	148

Next, the EI estimation models were built using subset X (PSBTI profile) as the training data set and subset Y (PSBT profile) after its imputation based on absolute average and relative % average. The regression estimation models for the datasets are –see Table 13:

- EI = 250hrs + 0.55xEP + 0.10xES -0.03xEB – 0.09xET(absolute EI imputed with an adjusted R^2 of 0.49).
- EI = 21hrs + 0.64xEP + 0.13xES -0.04xEB – 0.01xET(relative EI imputed with an adjusted R^2 of 0.69).

Both have a P-value <0.1 for both imputations which indicates the results are statistically significant at *t*-test and P-values for EI estimates.

Table 13: Simple imputations based on averages: EI estimation models (N=40)

Dataset	Intercept	Effort Plan	Effort Specify	Effort Build	Effort Test	Adjusted R^2	R^2	P-value
Imputation based on absolute EI average	**250**	**0.55**	**0.10**	**-0.03**	**-0.09**	**0.49**	**0.54**	**<0.0001**
Imputation based on relative % EI average	**21**	**0.64**	**0.13**	**-0.04**	**0.01**	**0.69**	**0.72**	**<0.0001**

The analysis of estimate variance was carried out after applying the estimation models built from subset X on subset Y – see Table 14.The quality of the estimation model for the imputation based on:

- Absolute EI average of subset X led to an MMRE = 155% and Pred(25) = 23%;
- Relative %EI average led to an MMRE = 74% and Pred(25) = 30%.

In summary, the imputation based on the relative EI average led to better estimation models of EI than those based on EI absolute average (i.e. adjusted R^2 = 0.69, MMRE = 74% and Pred(25) = 30%.

Table 14: Simple imputation: estimate variance of EI

No.	Dataset	N=40 projects, without outliers	
		%MMRE	Pred(25)
1	**Imputation based on absolute average**	**155%**	**23%**
2	**Imputation based on relative %average**	**74%**	**30%**

6.2 EI estimation with simple imputation and MI

This section presents the results of the various types of imputations from subset X as applied to subset Y based on:

- Simple imputation with the absolute average,
- Simple imputation with the relative % average,
- MI with values selected randomly from:
  - Absolute Min & Max EI seeds, and
  - Relative Min & Max EI seeds.

Table 15 presents the imputation results. For example, in comparison to the EI average of 6.7% for subset Y(from the actual values in Table 10), the imputed values in Table 13 vary from 8.2% using imputation based on relative % average of 8.2% to 19.0% when based on relative min and max seeds of 1% and 41%.

Table 15: Average effort distribution after imputation: for imputed subset Y(N=40)

Imputation based on	Profile	Project Phases – % Effort					No. of projects
		Effort Plan	Effort Specify	Effort Build	Effort Test	Effort Implement	
Absolute average	subset X: PSBTI	8.2	28.3	38.1	17.2	8.2	**20**
	subset Y: PSBT	8.6	23.4	39.3	18.2	10.6	**20**
Relative %average	subset Y: PSBT	11.1	17.0	48.6	23.4	8.2	**20**
MI absolute seeds	subset Y: PSBT	11.1	17.0	48.6	23.4	13.5	**20**
MI Relative seeds %1 to %24	subset Y: PSBT	11.1	17.0	48.6	23.4	13.0	**20**
MI Relative seeds %1 to %41	subset Y: PSBT	11.1	17.0	48.6	23.4	19.0	**20**

Table 16presents the average effort distribution by project phase for the full 40 projects, with subset X and the imputed subset Y recombined. The percentage effort distribution by phase for the full data set of 40 projects was 7.7%, as in Table 10 (e.g., the reference value for comparison purposes).

The relative EI average was used to compare the performance of the different types of imputations. Table 16shows that:

- For EI at 8.2%, the imputation based on the relative %average was the closest to the reference EI value of 7.7%.
- For EI at 11.7%, the imputation based on the relative seeds (1%, 41%) had the largest difference relative to the reference EI value of 7.7%.

Table 16: Average effort distribution after imputation: full set (N=40)

Imputation based on	Profile	Project Phases – % Effort					No. of projects
		Effort Plan	Effort Specify	Effort Build	Effort Test	Effort Implement	
Complete data	Subset X & Y: PSBTI	8.9	24.1	40.5	18.8	7.7	**40**
Absolute average	Subset X & Y: PSBTI	8.6	23.4	39.3	18.2	10.6	**40**
Relative %average	Subset X & Y: PSBTI	9.1	24.7	41.5	19.2	8.2	**40**
MI absolute seeds	Subset X & Y: PSBTI	9.1	24.7	41.5	19.2	9.9	**40**
Relative seeds %1 to %24	Subset X & Y: PSBTI	9.1	24.7	41.5	19.2	9.7	**40**
Relative seeds %1 to %41	Subset X & Y: PSBTI	9.1	24.7	41.5	19.2	11.7	**40**

7. Comparison of estimation performance from MI and simple imputations

This section presents a comparison of the MI estimation performance with the results from a simpler imputation technique based only on an average (absolute or relative). The EI estimation error variances of the set of 40 projects are compared in Tables 17 and 18 where the top half of the tables presents the adjusted R^2, %MMRE and Pred(25) for each model, while the bottom half of the tables presents pair-wise comparison of models performances.

7.1 Comparison between MI and simple imputation

For Option **A** (e.g., MI with absolute seeds):

a) The prediction performance with an MMRE = 116% and a Pred(25) = 27% (Table 3) were much better than the corresponding ones for the simple imputation with absolute average values (MMRE = 155% and Pred(25) =23% (Table 14). This corresponds to (line 6, Table 17):

- 16% increase in the adjusted R^2,
- 39% decrease in MMRE, and
- 4%increase in Pred(25).

In summary, with option A, when an EI estimation model is built through MI with absolute min-max seeds, we obtain a predictive accuracy that is statistically significantly better than from a simple imputation with an absolute average seed.

For option B(e.g., MI with relative seeds**)**:

a) Comparison of the MI based on relative seeds (1% to 24%) with respect to a simple imputation technique also based on a relative average corresponded to (line 7, Table 17):
   - 9% decrease in the adjusted R^2,
   - 73% increase in MMRE, and
   - 6% decrease in Pred(25).

b) Comparison of the MI based on relative seeds (1% to 41%) with respect to a simple imputation technique based also on a relative average corresponded to(line 8, Table 17):
   - 16% decrease in the adjusted R^2,
   - 98% increase in MMRE, and
   - 7% decrease of in Pred(25).

In summary, with option B, when an EI estimation model is built from simple imputed data based on a relative average, we obtain a predictive accuracy that is statistically significantly better than MI from with relative min-max seeds.

Table 17: Comparison of models predictive performances – relative vs absolute seeds (N=40)

No.	Dataset	Adjusted R^2	%MMRE	Pred(25)
1	**Imputation based on absolute average**	**0.49**	**155%**	**23%**
2	**Imputation based on relative %average**	**0.69**	**74%**	**30%**
3	**Imputed MI based on absolute seeds**	**0.65**	**116%**	**27%**
4	**Imputed MI (relative seeds %1 to %24)**	**0.60**	**147%**	**24%**
5	**Imputed MI (relative seeds %1 to %41)**	**0.53**	**172%**	**23%**
6	3 vs 1	**+16%**	**-39%**	**+4%**
7	4 vs 2	**-9%**	**-73%**	**-6%**
8	5 vs 2	**-16%**	**+98%**	**-7%**

7.2 Comparison with the reference data set (e.g., complete values)

The estimation model built from the 40 projects with data for the full PBSTI profile (presented in line 1, Table 1) is the reference in this study with an adjusted R^2 = 0.58, an MMRE = 88% and Pred(25) = 30% - Table 3). The estimation models built from MI and simple imputations were compared to this reference basis.

a) Imputation with absolute min-max seeds:

- For simple imputation(line 7, Table 18):
  - 22% decrease in the adjusted R^2,
  - 67%increase in MMRE, and
  - 7% decrease in Pred(25).

- For MI (line 9, Table 18):
    - 6% decrease in the adjusted R^2,
    - 28% increase in MMRE%, and
    - 3% increase in Pred(25).

b) Imputations with relative seeds:

- For simple imputation(line 8, Table 18):
    - 2% minor decrease in the adjusted R^2,
    - 14% increase in MMRE, and
    - No difference for Pred(25).
- For MI and relative seeds 1% to 24%(line 10, Table 18):
    - 11% decrease in the adjusted R^2,
    - 59% increase in MMRE, and
    - 7%decrease in Pred(25).
- For MI and relative seeds 1% to 4%1:
    - 18% decrease in the adjusted R^2,
    - 84%increasein MMRE, and
    - 7% decrease in Pred(25) (line 11 Table 18).

To summarize, no imputation technique is expected to perform better than the original dataset, of course, and as expected, with 50% missing data, the estimates from MI are not as good compared the estimates from a model trained with the actual data.

This is an encouraging result since this dataset was relatively small, N= 40 projects, and contained 50% missing observations, which are challenging circumstances for imputation techniques. Our empirical results suggest that the MI imputation method has practical utility for software engineers involved in effort estimation data analysis. In addition, it is worth observing that imputation is one of the activities in the more general field of data editing which includes a whole range of techniques for identifying, removing and updating suspect data.

Table 18: Comparison of models predictive performances vs complete dataset – reference (N=40)

No.	Dataset	Adjusted R^2	%MMRE	Pred(25)
1	**Complete dataset**	**0.71**	**88%**	**30%**
2	**Imputation based on absolute average**	**0.49**	**155%**	**23%**
3	**Imputationbased on relative %average**	**0.69**	**74%**	**30%**
4	**Imputed MI based on absolute seeds**	**0.65**	**116%**	**27%**
5	**Imputed MI (relative seeds %1 to %24)**	**0.60**	**147%**	**24%**
6	**Imputed MI (relative seeds %1 to %41)**	**0.53**	**172%**	**23%**
7	2 vs. 1	**-22%**	**+67%**	**-7%**
8	3 vs. 1	**-2%**	**+14%**	**0%**
9	4 vs. 1	**-6%**	**+28%**	**-3%**
10	5 vs. 1	**-11%**	**+59%**	**-6%**
11	6 vs. 1	**-18%**	**+84%**	**-7%**

8. Summary and discussion

The ISBSG repository comprises project data from several different organizations around the world. However, this repository contains a large number of missing data, which considerably reduces the number of data points available for building productivity and estimation models. There are a few techniques available for handling missing values, including simple and multiple imputations, but it is essential to apply them appropriately, otherwise biased or misleading inferences may be made.

The initial research goal of (Bala & Abran 2016) was to investigate only the use of the multiple imputation (MI) technique with the ISBSG repository for dealing with outliers and missing values. In this new study reported here we used both simple and MI techniques using either absolute and relative seeds, to build effort estimation models using regressions on the imputed datasets and then compared their respective performance.

In these studies, ISBSG R9 was used since subsequent ISBSG releases did not provide information on project phases. This dataset contained 41 projects with effort data for each of the ISBSG project phases, including one outlier on the independent variables. When this outlier was excluded for building the EI estimation models using multi-regression on the other project phases as independent variables for the 40 projects, the adjusted R^2 improved from 0.58 (Table 1) to 0.71 (Table 2). Therefore, removing the outlier strengthened the linearity of the data and decreased the errors present in the regression. Furthermore, the results were statistically significant for the estimates of EI, as illustrated by the t-test and P-values without the outlier.

The performance of the MI and simple imputation techniques were investigated using the set of 40 projects in the ISBSG dataset with data in their PSBTI profile and dividing it into 2 subsets: subset X with 20 projects with complete data fields and subset Y with the other 20 projects from which the information in the EI data field was deleted (e.g., thereby corresponding to a PSBT profile). Subset X of the complete data of PSBTI was used as the training dataset for the MI and for building the estimation model, and then this estimation model was applied to subset Y with EI missing values. The performance of the effort prediction models between estimated effort and actual effort was evaluated using three evaluation criteria: Adjusted R^2, MMRE, and Pred(25). The estimates were obtained from multiple regression analysis estimation models.

Two distinct options were designed to investigate MI impact:

- Option A: Using MI with seeds from the absolute min-max values of EI – this was referred to as imputation from absolute EI values. Results of option A were presented in Section 4.
- Option B: Using MI with seeds from relative min-max values of EI values (e.g., when the seeds are the min-max values of EI relative to Total Effort. Results of option B were presented in Section 5.

Next, to compare the performance of MI, Section 6 replicated the study of (Déry & Abran 2005)] with simple imputations with absolute average. In addition, simple imputation with relative EI average was investigated.

Findings from MI for both options A and B and from simple imputations based on either the absolute or the relative seeds were compared in section 7. The key findings with the data set used in this study and deletion of 50% of the data in the EI field are:

- When only absolute value was calculated: MI with absolute seeds (line 3, Table 17) performed much better in comparison to simple imputation with an absolute average.
- When a value of the variable being estimated (here EI) was calculated relative to other fields (here, the effort for each of the project phases): then, simple imputation with a relative average seed (line 2, Table 17) performed best for all techniques and types of seeds tested, and better than MI with either absolute or relative seeds.
- When estimates were compared to the complete dataset (e.g., reference dataset here – line 1, Table 18):

- Simple imputation with a relative average seed (line 3, Table 18) performed best for all techniques and types of seeds tested, with estimation results almost as good as estimates from the complete dataset on all three criteria of Adjusted R^2, MMRE, and Pred(25).
- MI with absolute seeds (line 4, Table 18) was the next best strategy with close Adjusted R^2 and Pred(25), but degraded MMRE. This was the alternative when other variables were not available to calculate relative seeds.

Our empirical results suggest that MI techniques have practical utility for software engineers involved in effort estimation data analysis.

It is worth noting that imputation is just one activity in the general field of data editing, which includes a whole range of techniques for identifying, removing and updating suspect data. Our research methodology and results provide practical and substantiated guidelines for researchers and practitioners constructing effort estimation models when their datasets have outliers and missing values. Additional related research should be pursued to derive more general results. For example, researchers may develop new techniques, or improve existing techniques, to investigate fields with a large number of missing values in the ISBSG data repository, such as max team size, lines of code, and resource level.

MI is not the only modern missing data tool available to researchers. Some producers of statistical software are beginning to incorporate incomplete data features directly into certain types of modeling routines. These procedures are similar to MI in that they implicitly average over a predictive distribution for the missing values, but the averaging is performed using analytic or numerical methods rather than simulation.

This research encourages future replications of simulations on the ISBSG dataset reported in this empirical study in order to confirm our conclusions. Certainly, such replications will have important practical significance for practitioners and researchers building effort estimation models.

References

Abran, Alain (2015) Estimation of Software Projects, IEEE-CS Press & John Wiley & Sons, p. 328.

Acuña E. & Rodriguez C. (2004) 'The Treatment of Missing Values and its Effect on Classifier Accuracy', in: Classification, Clustering, and Data Mining Applications - Studies in Classification, Data Analysis, and Knowledge Organisation, Springer, Berlin, Heidelberg, pp. 639-648.

Bala, Abdalla & Abran, Alain (2016) 'Use of the Multiple Imputation Strategy to Deal with Missing Data in the ISBSG Repository', Journal of Information Technology & Software Engineering,Vol. 6, No.1, p. 12.

Cartwright, MH., Shepperd, MJ. and Song, Q. (2003) 'Dealing with missing software project data',IEEE Software Metrics Symposium, Sydney, Australia,pp. 154-165.

Conte, SD.& Dunsmore, HE. & Shen, VY. (1986) Software Engineering Metrics and Models,Benjamin/ Cummings Publishing Company Inc., p. 403.

Deng, K. & MacDonell, SG. (2008) 'Maximising data retention from the ISBSG repository', 12th International Conference on Evaluation and Assessment in Software Engineering - EASE, Bari, Italy, pp. 21-30.

Déry, David & Abran, Alain (2005) 'Investigation of the Effort Data Consistency in the ISBSG Repository', 15th International Workshop on Software Measurement - IWSM, Montreal, Canada, Shaker-Verlag, pp. 353-368.

Graham, JW. & Schafer, JL. (1999) 'On the performance of multiple imputation for multivariate data with small sample size', in Statistical strategies for small sample research,Hoyle R (ed.),pp.1-29.

Horton, N. & Lipsitz, S. (2001) 'Multiple Imputation in Practice: Comparison of Software Packages for Regression Models with Missing Variables',The American Statistician, Vol. 55, No. 3, pp. 244-254.

Idri, Ali & Abnane, Ibtissam & Abran, Alain (2016) 'Dealing with Missing Values in Software Project Datasets: A Systematic Mapping Study', book chapter in Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, in Studies in Computational Intelligence, No. 653, Springer, pp. 1-16.

ISBSG, International Software Benchmarking Standards Group, Analysis Reports, http://www.isbsg.com/collections/analysis-reports

Little, R. (1998) 'A Test of Missing Completely at Random for Multivariate Data with Missing Values',Dealing with missing software project dataJournal of the American Statistical Association, Vol. 83, No. 404, pp. 1198-1202.

Little, R. & Rubin, D. (2002) Statistical Analysis with Missing Data, 2nd Edition, Wiley-Interscience.

Magnani, M. (2004) Techniques for Dealing with Missing Data in Knowledge Discovery Tasks, Department of computer Science, University of Bologna,Italy.

Mockus, A. (2008) Missing Data in Software Engineering, Springer, London.

Myrtveit, I. & Stensrud, E. & Olsson, UH. (2001) 'Analyzing data sets with missing data: an empirical evaluation of imputation methods and likelihood-based methods', IEEE Transactions on Software Engineering, Vol. 27, pp. 999-1013.

Rubin, D. (1987) Multiple Imputation for Nonresponse in Surveys,Wiley Series in Probability and Statistics, John Wiley & Sons.

Song, Q. & Shepperd, Martin (2007) 'A new imputation method for small software project data sets',Journal of Systems and Software, Vol. No. 1, pp. 51-62.

Strike, K. & El Eman, Khaled & Madhavji, Nazim (2001) 'Software cost estimation with incomplete data',IEEE Transaction on Software Engineering, Vol. 27, No. 10, pp. 890-908.

Empirische Untersuchung zum Einsatz von SaaS-Lösungen bei Klein- und mittelständischen Unternehmen (KMUs)

Andreas Schmietendorf, André Nitze

Hochschule für Wirtschaft und Recht Berlin

Email: andreas.schmietendorf@hwr-berlin.de

1. Ziele der empirischen Untersuchung

Das Cloud-Computing hat die Art und Weise, wie IT-Leistungen durch Anbieter bereitgestellt bzw. durch Nutzer verwendet werden, gravierend verändert. Die zumeist browserbasierte Nutzung von Infrastrukturen, Plattformen oder ganzen Softwarelösungen ist sowohl im privaten als auch im industriellen Diskurs weithin akzeptiert. Mit diesen internetbasierten Servicemodellen gehen vielfältige technische, organisatorische, sicherheitsrelevante, wirtschaftliche und rechtliche Herausforderungen einher. Für Hintergründe zum Cloud-Paradigma sei auf entsprechende Literatur wie z.B. [Höllwarth 2013], [Vossen 2013] oder auch [Terplan 2011] verwiesen.

Gerade bei Klein- und Mittelständischen Unternehmen (kurz KMU) lässt sich ein differenziertes Verhalten bezüglich der Art eingesetzter Softwarelösungen beobachten. Mit Hilfe einer empirischen Analyse sollte u.a. auf folgende Fragen eingegangen werden:

- Was sind die Treiber zur Nutzung von SaaS (Software as a Service)?
- Welche Chancen und Risiken sehen die Unternehmer in SaaS?
- Welche Veränderungen lassen sich im zeitlichen Verlauf beobachten?
- Wie wird mit Qualitätseigenschaften durch Anbieter/Nutzer umgegangen?
- Wo werden Hemmnisse in Bezug auf den SaaS-Einsatz gesehen?
- Welche Arten von Anwendungen werden mit Hilfe von SaaS-Services bezogen?
- Wie beeinflusst eine zunehmend mobil orientierte Arbeitswelt die SaaS-Nutzung?

Eine Umfrage bei KMUs zur Verwendung von SaaS-Angeboten wurde erstmals im Jahr 2015 und abermals im Jahr 2017 durchgeführt. Der vorliegende Beitrag konzentriert sich auf eine vergleichende Darstellung der Ergebnisse beider Umfragen.

2. Existierende Arbeiten

Die meisten Arbeiten zu empirisch gewonnenen Ergebnissen wurden im industriellen Diskurs erarbeitet. Wissenschaftliche Publikationen haben sich lange mit den grundlegenden Eigenschaften des Cloud-Paradigmas, den Möglichkeiten einer modellbasierten Abbildung oder auch den betrieblich bedingten Implikationen eines Cloud-Einsatzes auseinander gesetzt. Im Folgenden sollen bewusst Arbeiten charakterisiert werden, welche sich auf einen empirischen Erfahrungshintergrund beziehen. Dabei kann selbstverständlich kein Anspruch auf Vollständigkeit erhoben werden.

Bereits im Jahr 2011 wurde durch das BMWT ein Leitfaden zum Cloud Computing für den Bereich der mittelständischen Unternehmen zur Verfügung gestellt. (vgl. [BMWT 2011]) Entsprechend dargestellte Kurzstories orientierten sich an realen Anwendungsszenarien, ebenso wird eine Vorschau auf erwartete Ergebnisse (z.B. Anzahl von Arbeitsplätzen durch Cloud Computing) im Jahr 2015 gegeben.

Unter [Hamburg/Bucksch 2016] findet sich ein Beitrag zu den Ergebnissen einer Befragung von 260 KMUs aus dem Jahr 2015. Neben einer begrifflichen Einordnung erfolgt die Darstellung ausgewählter Ergebnisse zur den Gründen einer Nichtnutzung bzw. zu den erwarteten Vorteilen von Cloud-Lösungen.

Die bereits unter [Schmietendorf 2016] genannten Untersuchungen der Experton Group (Cloud-Vendor Benchmark) und der KPMG/Bitkom (Cloud-Monitor) haben eine entsprechende Fortschreibung erfahren. (vgl. [Experton 2016] bzw. [Pols/Vogel 2017]). Vergleichbare Studien finden sich im internationalen Kontext, wie z.B. unter [RightScale 2017] oder [BCSG 2014].

Mit der Durchführung einer eigenen Studie sollte dem speziellen Fokus auf KMUs aus dem deutschsprachigen Raum und der expliziten Berücksichtigung differenzierter SaaS-Angebote Rechnung getragen werden. Darüber hinaus erfolgte in vielen Studien keine ausreichende Differenzierung zwischen SaaS-Anbieter- und –Nutzer bzw. der expliziten Erfassung von Nicht-Nutzern und deren speziellen Rahmenbedingungen.

3. Konzept der Forschungsarbeit

Die empirische Bewertung der Akzeptanz von SaaS-Lösungen wurde mit Hilfe eines mehrstufigen Forschungsansatzes herausgearbeitet. Dafür erfolgte eine enge Zusammenarbeit mit verschiedenen Praxispartnern aus dem KMU-Umfeld. Im Kern handelt es sich um die folgenden Themenbereiche, deren Teilergebnisse im Rahmen vorangegangener WIWITA-Tagungen bereits publiziert wurden:

- Entwicklung allgemeiner Bewertungs- und Messansätze für SaaS-Lösungen, welche zur Serviceauswahl und Gestaltung von Verträgen herangezogen werden können (vgl. [Schmietendorf 2012])

- Neben einer umfangreichen Quellenrecherche galt die Aufmerksamkeit der Analyse konkreter Einführungsprojekte von Cloud- bzw. SaaS-Service bei mittelständischen Unternehmen. (vgl. [Schmietendorf 2014])

- Der Entwicklung der internetbasierten Umfrage, deren prototypischer und konkreter Verwendung sowie den dabei erreichten Ergebnissen widmete sich der nächste Forschungsschritt. (vgl. [Schmietendorf 2016] bzw. [Schmietendorf/Kunisch 2016])

Im Jahr 2017 wurde die internetbasierte Umfrage abermals durchgeführt, wofür im Vorfeld eine leichte Anpassung der verwendeten Fragen erfolgte. Der Fokus dieser Überarbeitung bezog sich auf die Gewährleistung der Vergleichbarkeit, die Beseitigung von während der ersten Durchführung erkannten Schwächen, die Berücksichtigung von aktuellen Entwicklungstendenzen und die Anpassung auf die aktuelle Version des verwendeten Systems zur Implementierung und Auswertung der Umfrage (vgl. https://de.surveymonkey.com).

Die Strukturierung des Fragebogens erfolgte in der folgenden Weise:

- 5 Fragen für Nicht-Nutzer (4 Fragen in 2015)

- 14 Fragen für Nutzer (11 Fragen in 2015)

- 10 Fragen für Anbieter (9 Fragen in 2015)

- 5 Fragen zum unternehmerischen Diskurs.

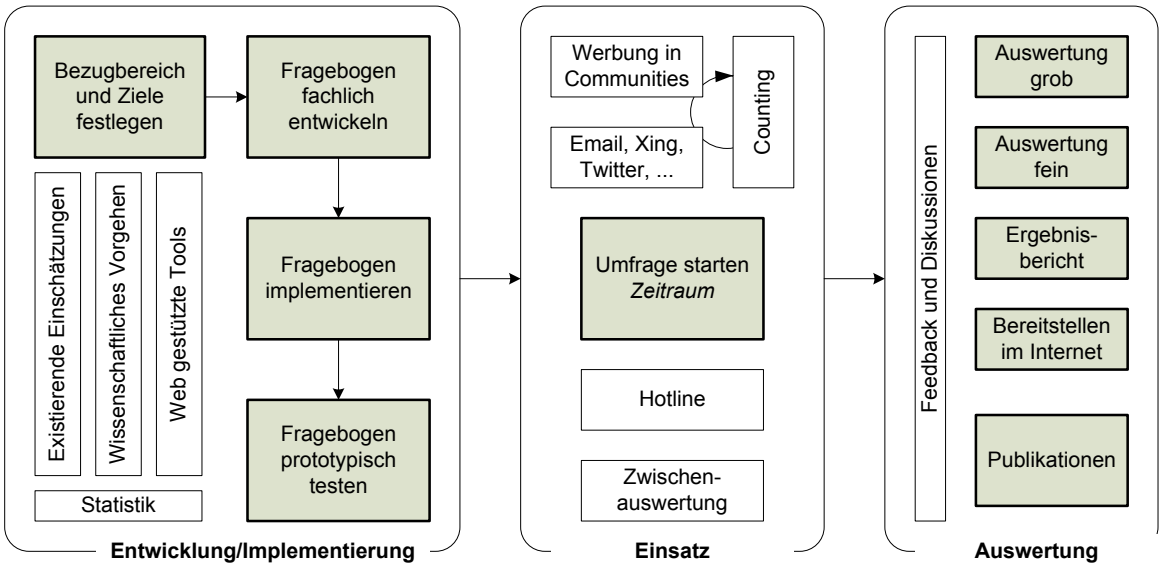


Abbildung 1: Konzept der durchgeführten Umfrage [Schmietendorf 2016]

Der bereits entwickelte Prozess zur Entwicklung, Durchführung und Auswertung der Umfrage wurde nicht verändert. Dessen abermalige (interative) Verwendung erwies sich als hilfreich für die Strukturierung, Planung und Überwachung der durchzuführenden Aufgaben. Als erfolgskritisch erwies sich abermals die Bewerbung der Umfrage. Zum erreichen eines statistisch relevanten Ergebnisses galt es Führungskräften aus Klein- und mittelständischen Unternehmen für die Teilnahme an der Umfrage zu gewinnen.

4. Ergebnisse der Untersuchung

Im Rahmen dieses Beitrags kann nur auf ausgewählte Ergebnisse der Umfrage eingegangen werden. In der Anlage findet sich die grafisch aufbereitete Darstellung der wichtigsten Ergebnisse im Überblick. Der komplette Ergebnisbericht kann über die folgende Adresse im Internet abgerufen werden ([Schmietendorf/Kunisch 2018]): https://blog.hwr-berlin.de/schmietendorf/)

4.1 Summarische Darstellung

An der Umfrage haben sich insgesamt 142 Personen (43% Nutzer, 32% Anbieter) beteiligt. Bezüglich der vertretenen Branchen gab es eine Beteiligung aus der Informationsverarbeitung, der Industrie, der Beratung, dem öffentlichen Dienst, dem Maschinenbau, aus Banken/Versicherungen, der Energieversorgung, dem Gesundheitswesen, dem Handel, Chemie/Pharma und der Bauwirtschaft.

- Noch vor zwei Jahren war sich jeder fünfte Befragte gar nicht sicher, ob in seinem oder ihrem Unternehmen überhaupt SaaS-Services eingesetzt wurden. Heute haben rund vier von fünf Befragten (79%) Software aus der Cloud im Unternehmenseinsatz oder bieten sie selbst an. Zwei Drittel der Befragten sind Nutzer. Diese Entwicklung spricht für eine starke Durchdringung von Cloud-Angeboten im Mittelstand und ihre noch immer zunehmende Bedeutung.

- Bei den Befragten, die angaben, gar keine SaaS-Lösung zu nutzen (21%), war der Grund wieder einhellig die Sorge um Datensicherheit und das Problem, zwar Daten übertragen zu können, aber nicht die Verantwortung dafür zu haben. Hoher Einführungs- bzw. Integrationssaufwand sowie schlechte Performance sind weitere Argumente der Cloud-Gegner.

- Die Teilnehmer, in deren Unternehmen SaaS-Lösungen eingesetzt werden, bewerten ihre Anbieter überwiegend positiv. Zwei Drittel gaben Ihren Anbietern die Schulnote „2“. Am meisten schätzen die Anwender neben den eigentlichen Produktfunktionen dabei vor allem die bedarfsgerechte Nutzung und Abrechnung der Software sowie den professionellen Support.

- Ein Drittel der Befragten gab an, dass mehr als fünf verschiedene SaaS-Angebote in ihrem Unternehmen eingesetzt werden, die auch noch so gut wie alle (91%) von verschiedenen Anbietern stammen. Um nicht den Anschluss zu verlieren, legen 72% der Anwender entsprechend großen oder sehr großen Wert auf die Integrationsfähigkeit ihrer Cloud-Software. Standardisierte Schnittstellen (APIs) sind der Schlüssel zur Interoperabilität zwischen SaaS-Lösungen.

- Der eindeutige Trend: Cloud-Speicher ist als IT-Commodity auf breiter Front (68%) angekommen. Die Hälfte der befragten Nutzer setzt in ihren Unternehmen zudem Cloud-basierte Kommunikationsdienste ein. Danach kommen mit jeweils ungefähr einem Drittel Durchdringung typische Unternehmensanwendungen wie ERP, Content- und Dokumenten-Management sowie Projekt- und Personalverwaltung.

- Vertrieb, Marketing und Verwaltung sind Vorreiter bei der Nutzung von Cloud Computing – in jeweils einem Drittel der Unternehmen sind SaaS-Lösungen in diesen Bereichen bereits im Einsatz. Mittlerweile nimmt auch die Nutzung von SaaS in Personalwesen und Buchhaltung zu: Laut Befragung nutzen derzeit 23 Prozent der Buchhaltungen und 27 Prozent der Personalabteilungen SaaS, Tendenz weiter steigend.

- Cloud-Marktplätze haben - wie auch in der ersten Befragung - kaum praktische Relevanz für die Nutzer. Obwohl bereits die Hälfte der Hersteller ihre Produkte auf solchen Marktplätzen anbietet oder dies plant, beziehen so gut wie alle Kunden ihre SaaS-Lösungen ohne den Weg über die Marktplätze. Offenbar nehmen Kunden hier noch keinen Mehrwert wahr bzw. werden von bekannten IT-Partnern seht gut betreut.

4.2 Herausforderungen und Gründe für SaaS-Lösungen

Im Folgenden sollen zwei Ergebnisse der Nutzerbefragung dargestellt werden. Zunächst galt das Interesse der Frage, worin Nutzer anstehende Herausforderungen bei SaaS-Services sehen:

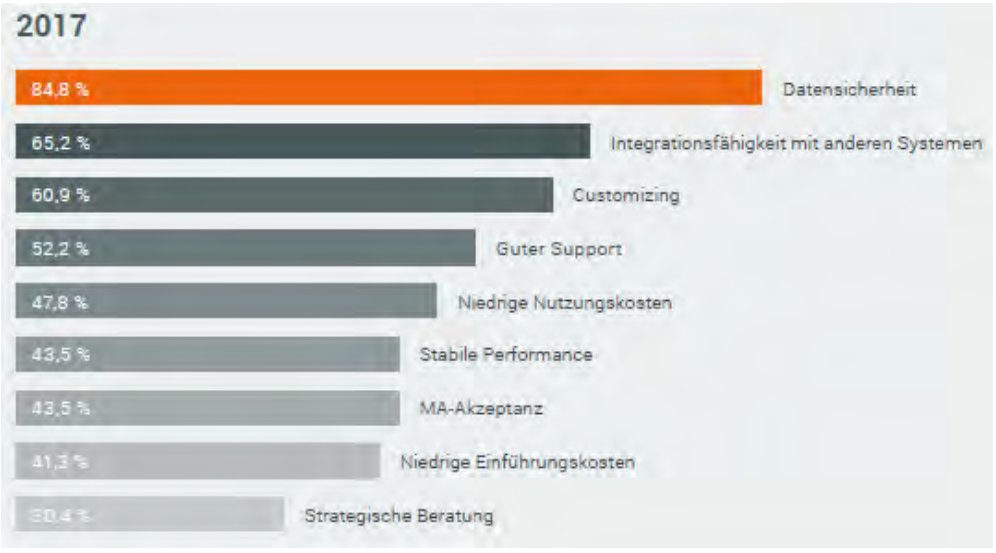


Abbildung 2: **Herausforderungen für SaaS aus Nutzersicht [Schmietendorf/Kunisch 2018]**

Im Vergleich zum Jahr 2015 hat sich an der Reihung kaum etwas verändert. Allerdings haben die Kriterien *niedrige Nutzungskosten*, *Mitarbeiterakzeptanz*, *niedrige Einführungskosten* und *strategische Beratung* den stärksten Zuwachs erfahren.

Die Möglichkeit des mobilen Arbeitens ist aktuell der beliebteste Grund für die Einführung Cloud-basierter Software. Offenbar passt dieser SaaS-Mehrwert gut zu den sich verändernden Arbeitsgewohnheiten. Fast genauso wichtig erschienen den Befragten jedoch auch wieder Kosteneffizienz, Entlastung der IT-Abteilung und Skalierbarkeit. Die Konzentration auf das Kerngeschäft wird als deutlich reduzierter Grund für die SaaS-Einführung angesehen. Interessant ist auch die veränderte Bewertung der über SaaS-Lösungen erfahrenen Benutzerfreundlichkeit, offensichtlich wurden hier positive Erfahrungen gemacht.

Abbildung 3: Vergleich der Gründe für SaaS aus Nutzersicht [Schmietendorf/Kunisch 2018]

4.3 Thesen zur Fortentwicklung

These 1

„Software-as-a-Service“ wird zu „Software“ bzw. „Cloud Computing“ zu „Computing“

Die Differenzierung zwischen SaaS- und On-Premise-Lösungen verschwimmt für den Anwender zunehmend. Aufgrund der „black box“-Sicht auf die hinter einer SaaS-Lösung liegenden Leistungserbringer kann die Speicherung und Verarbeitung von Daten durch den Anwender nicht mehr nachvollzogen werden. Dem entsprechend gilt die Aufmerksamkeit vertraglich vereinbarter Qualitäts- und Leistungsmerkmale, was sich so auch in den Befragungsergebnissen niederschlägt. Damit wird aus „Software-as-a-Service“ einfach nur noch „Software“.

These 2:

Mobiles Arbeiten ist eine Grundvoraussetzung für eine zunehmend digitalisierte Arbeitswelt

Eine der größten Veränderungen in der diesjährigen Befragung ist die enorm gestiegene Bedeutung des mobilen Arbeitens als Argument für den Einsatz von SaaS-Lösungen. Mitarbeiter möchten bei Bedarf flexibel von zu Hause oder von unterwegs aus arbeiten und nutzen dabei sowohl verschiedene Geräte (Rechner, Smartphone, Tablet …) als auch unterschiedliche Datencontainer (Google Drive, Dropbox oder Sharepoint). Ein Mitarbeiter hat sehr wahrscheinlich mehrere digitale Identitäten (AD Account, Google ID, Apple ID) und arbeitet mit unterschiedlichsten Anwendungen.

These 3:

Cloud-basierte Service-APIs sind Enabler der Digitalisierung

Im Mittelpunkt des aktuell verwendeten Digitalisierungsbegriffs steht die Integration von bisher disjunkt betrachteten Softwarelösungen. Entsprechende Anwendungen (beispielsweise Industrie 4.0) überspringen Unternehmens-, Branchen- und Technologiegrenzen. Cloudbasierte Programmierschnittstellen (APIs) sind der Dreh- und Angelpunkt, um dieses Innovationspotenzial zu heben.

5. Zusammenfassung und Ausblick

Zusammenfassend lässt sich eine zunehmende Akzeptanz von SaaS-Lösungen im KMU-Umfeld feststellen. Den „Schritt in die Cloud“ gehen die meisten Unternehmen allerdings noch immer über zunächst unternehmensunkritische Bereiche, wie Speicher- und Kommunikationslösungen oder auch Dokumenten- bzw. Content-Management-Lösungen. Bei den befragten KMUs zeigt sich darüber hinaus eine hohe Zufriedenheit mit längerfristig bekannten IT-Partnern, welche Cloud-Lösungen entlang der speziellen Bedürfnisse ihrer Kunden zur Verfügung stellen und die Integration bzw. Migration entsprechend begleiten können. Damit lässt sich auch das geringe Interesse an entsprechenden Marktplätzen für SaaS-Angebote erklären. Inwieweit sich dieser Trend fortsetzt, bleibt abzuwarten, zumal Unternehmen, wie z.B. die Gartner Group, in entsprechende Verzeichnisdienste/Marktplätze (vgl. Capterra - https://www.capterra.com) verstärkt investieren.

Noch zögerlich ist die Verwendung von SaaS-Lösungen in stark prozessorientierten Bereichen wie beim CRM, beim ERP- oder gar beim SCM zu beobachten. Im Zusammenhang mit diesen Lösungen wären weiterführende Gründe/Bedenken, wie z.B. die Ablösung/Integration existierender Systeme, Sicherheitsbedenken, Abhängigkeiten von einem Anbieter oder auch die ggf. vorhandene Gefahr einer Übernahme von einträglichen Geschäftsmodellen, durch den SaaS-Anbieter zu hinterfragen.

Gerade im Diskurs einer zunehmenden Digitalisierung sind KMUs gefordert, sich auf ihr Kerngeschäft zu konzentrieren. Für alle Unternehmen steht sich hier die Frage nach der Teilhabe an innovativen Geschäftsmodellen, wofür Daten und Funktionen zu konsumieren, aber eben auch bereitzustellen sind. Sehr wahrscheinlich ist die parallele Ausprägung der Rollen eines Cloud-Nutzers aber auch der eines Cloud-Anbieters. Selbst klassische Unternehmen, wie z.B. John Deere im Bereich der Landmaschinentechnik, erkennen die Möglichkeiten von API-basierten Cloudangeboten, um so z.B. sensorisch erfasste Daten bereitzustellen.

6. Quellen

[BCSG 2014] The small business revolution: trends in SMB cloud adoption, https://www.bcsg.com/wp-content/uploads/2015/03/The-small-business-revolution-trends-in-SMB-cloud-adoption.pdf, last Upload März 2018

[BMWT 2011] Cloud Computing – Leitfaden für mittelständische Unternehmen, Nationaler IT Gipfel, München 2011

[Experton 2016] Cloud Vendor Benchmark 2016, Experton Group, 2016, http://research.isg-one.de/research/studien/cloud-vendor-benchmark-2016/overview.html

[Hamburg/Bucksch 2016] Hamburg, I.; Bucksch, S.: Cloud Computing in SMEs, In Proc.: ICIW 2016: The Eleventh International Conference on Internet and Web Applications and Services, IARIA, 2016

[Höllwarth 2013] Höllwarth, T.: (Herausgeber) Cloud Migration: Alles, was Sie über die Cloud wissen müssen, MITP Verlags GmbH & Co. KG, Oktober 2013

[Pols/Vogel 2017] Pols, A.; Vogel, M.: Cloud Monitor 2017 - Eine Studie von Bitkom Research im Auftrag von KPMG, Pressekonferenz, www.kpmg.de/cloud, März 2017

[RightScale 2017] State of the Cloud Report, https://www.rightscale.com/lp/2017-state-of-the-cloud-report, 2017

[Schmietendorf 2012] Schmietendorf, A.: Spezifikation, Analyse und Bewertung der Qualitätseigenschaften im Internet angebotener Cloud-Services, in Cleve, J. (Hrsg.): Proc. 8. Wismarer Wirtschaftsinformatik-Tage (WIWITA 2012), S. 51-60, Mai/Juni 2012

[Schmietendorf 2016] Schmietendorf, A.; Nitze, A.; Hentschel, J.: Nutzung von SaaS-Lösungen bei Klein- und mittelständischen Unternehmen, in Cleve, J.; Alde, E.: Proceedings 10. Wismarer Wirtschaftsinformatiktage (WIWITA 2016), S. 79-88, Hochschule Wismar 2016,

[Schmietendorf/Bauer 2014] Schmietendorf, A.; Bauer, M.: Erfahrungen bei der Einführung von prozessbezogenen Cloud-Services in mittelständischen Unternehmen, in Cleve, J.; Alde, E.: Proceedings 9. Wismarer Wirtschaftsinformatiktage (WIWITA 2014), S. 1-13, Hochschule Wismar 2014

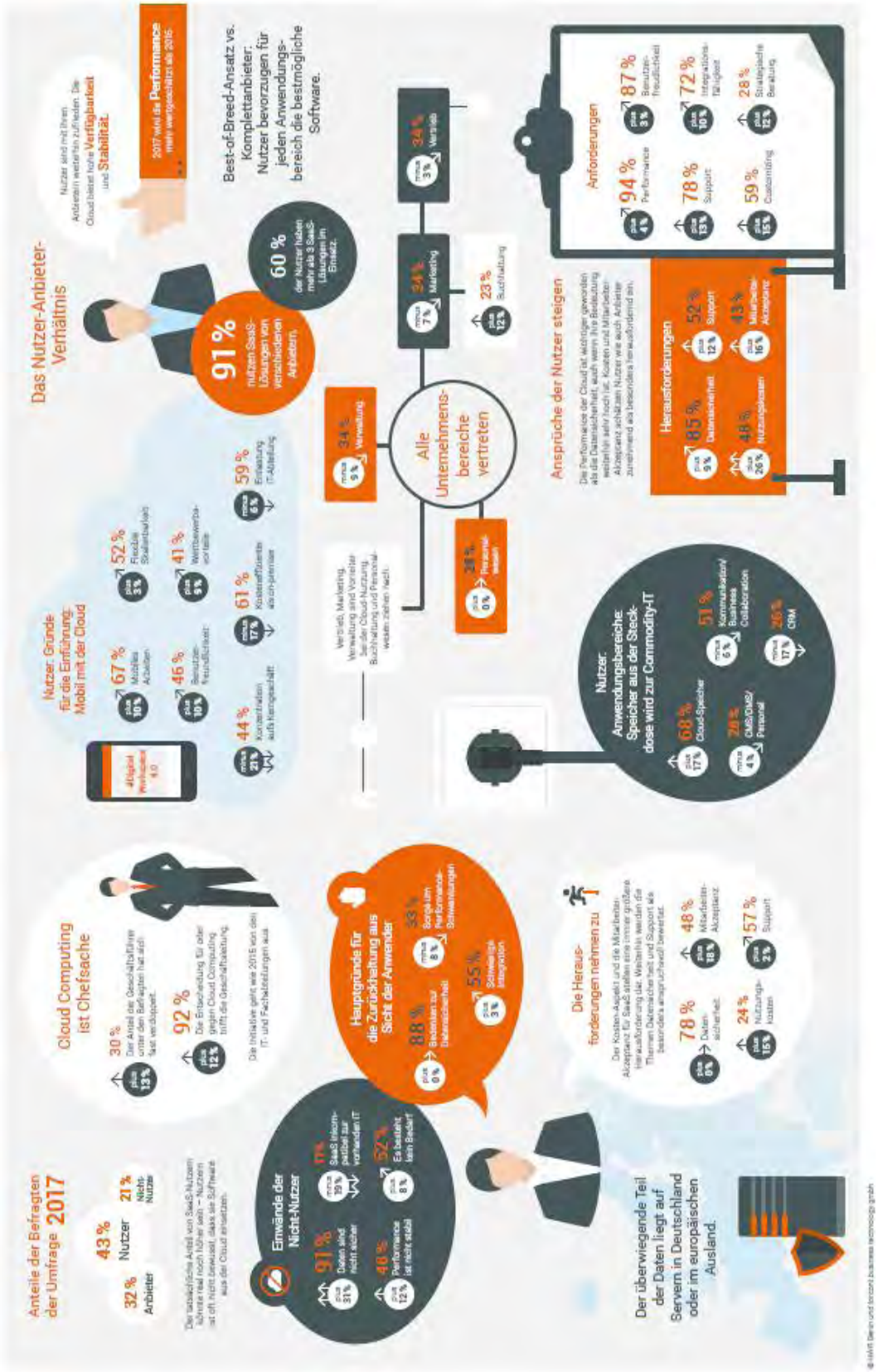
[Schmietendorf/Kunisch 2016] Schmietendorf, A.; Kunisch, M. (Hrsg.): Sind deutsche Unternehmen bereit für Cloud Computing? Eine Umfrage unter Anwendern und Anbietern von Software-as-a-Service, Redaktion/Layout: Möller Horcher Public Relations GmbH/Der Punkt GmbH, Februar 2016

[Schmietendorf/Kunisch 2018] Schmietendorf, A.; Kunisch, M. (Hrsg.): Deutscher Mittelstand auf Wolke 7? – Cloud Computing in kleineren und mittleren Unternehmen 2015 und 2017, Redaktion/Layout: Möller Horcher Public Relations GmbH/Der Punkt GmbH, Januar 2018, Studie abrufbar über: https://blog.hwr-berlin.de/schmietendorf

[Terplan 2011] Terplan, K.; Voigt, C.: Cloud Computing, MITP Verlags GmbH & Co. KG, 2011

[Vossen 2013] Vossen, G.; haselmann, T.; Hoeren, T.: Cloud-Computing für Unternehmen: Technische, wirtschaftliche, rechtliche und organisatorische Aspekte, dpunkt.verlag, 2013

Anlage: Überblick der Umfrageergebnisse



Quelle der Abbildung: [Schmietendorf/Kunisch 2018]

Dank

Der Dank gilt allen Autoren der Studie: André Nitze, OvG Universität Magdeburg; Andreas Schmietendorf, HWR Berlin/OvG Universität Magdeburg; Konrad Nadobny, HWR Berlin; Matthias Kunisch, forcont business technology. Ebenso gilt es, der „Möller Horcher Public Relations GmbH“ (speziell: Ines Bilas, Katja Dreißig) für die redaktionelle Bearbeitung zu danken sowie „Der Punkt GmbH“ für das Layout. Darüber hinaus gilt der Dank für die Unterstützung durch den Arbeitskreis Software-Qualität und -Fortbildung e. V. (ASQF), der Central Europe Computer Measurement Group e. V. (ceCMG) und der Deutschsprachige Anwendergruppe für Software-Metrik und Aufwandschätzung e. V. (DASMA).

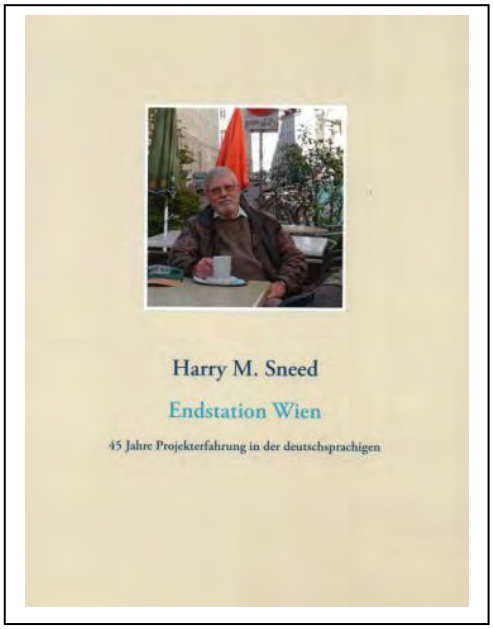
Harry Sneed:

Endstation Wien

45 Jahre Projekterfahrungen in der deutschsprachigen IT-Welt

BoD Norderstedt, 2017, 328 S.
ISBN 978-3-7448-8364-1

Dieses Buch beschreibt nahezu die gesamte Tätigkeit von Harry Sneed in der IT-Welt, von den Anfängen der Großrechner mit den COBOL und PL/1-Programmen bis hin zu den aktuellen und modernen Ansätzen Service-orientierter Technologien und Systemen. Dieses Buch fasst vor allem die umfangreichen Erfahrungen zu Wartungs-, Migrations- und Testprojekten zusammen, die auch für die Beherrschung aktueller und moderner Software-Anwendungen, wie beispielsweise autonomes Fahren oder Smart Cities, von unschätzbarem Wert sind. Es zeigt in unterhaltener Weise die Probleme bei der Software-Evolution und die teilweise immensen Anstrengungen für die kontinuierliche Gewährleistung von IT-Diensten.



Staron, M, Melding, W.:

Proceedings of the IWSM/Mensura 2017

Joined Conference of the 27th International Workshop on Software Measurement (IWSM) and the 12th International Conference on Software Process and Product Measurement (Mensura), ACM 2017, *ISBN 978-1-4503-4853-9*

This proceedings are available at the Computer Science Bibliography of Trier.

Capers Jones:

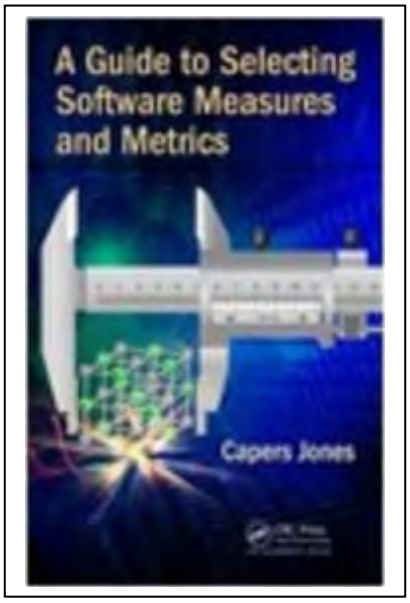
Guide to Selecting Software Measures and Metrics

Auerbach Publications, 2017, 358 p.
ISBN 978-1138-033078

The book helps software project managers and developers uncover errors in measurements so they can develop meaningful benchmarks to estimate software development efforts. It examines variations in a number of areas that include:

- Programming languages
- Development methodology
- Software reuse
- Functional and nonfunctional requirements
- Industry type
- Team size and experience

Filled with tables and charts, this book is a starting point for making measurements that reflect current software development practices and realities to arrive at meaningful benchmarks to guide successful software projects.



Jasveer Singh:

Functional Software Size Measurement Methodology with Effort Estimation and Performance Indication

John Wiley/IEEE Press 2017

This book describes a new size measurement method called as FSSM (as acronym from the book title) in order to qualify the effort and performance estimation of IT project. This method is addressed for software written in high-level languages and should useable for most of the current types of projects.

Schmietendorf, A.:

ESAPI 2017

1. Workshop: Evaluation of Service-APIs
2. November 2017, Berlin

Shaker Verlag, Aachen, 2017 (72 Seiten), ISBN 978-3-8440-5559-7

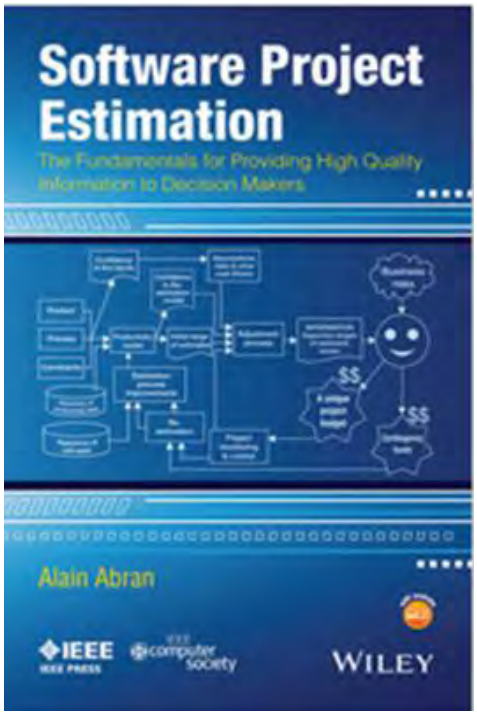
The book includes the proceedings of the first ESAPI 2017 Workshop held in Berlin in November 2017, which constitute a collection of theoretical studies in the field of measurement and evaluation of service oriented and API technologies.

Abran, A.:

Software Project Estimation: The Fundamentals for Providing

High Quality Information to Decision Makers

Wiley IEEE Computer Society Press, 2015 (288 pages), ISBN 978-1-118-95408-9



This book introduces theoretical concepts to explain the fundamentals of the design and evaluation of software estimation models. It provides software professionals with vital information on the best software management software out there.

- End-of-chapter exercises
- Over 100 figures illustrating the concepts presented throughout the book
- Examples incorporated with industry data

Heidrich, J.; Vogelezang, F.:

IWSM/Mensura 2016

Joined Conference of the 26th International Workshop on Software Measurement (IWSM) and the 11th International Conference on Software Process and Product Measurement (Mensura), IEEE Computer Society, CPS, http://www.computer.org/cps, 2016

This proceedings are available at the IEEE online publishing service.

Seufert, M.; Ebert, C, Fehlmann, T.; Pechlivanidis, S.; Dumke, R. R.:

MetriKon 2015 - Praxis der Softwaremessung

Tagungsband des DASMA Software Metrik Kongresses
5. - 6. November 2015, IBM, Köln

Shaker Verlag, Aachen, 2015 (272 Seiten)

The book includes the proceedings of the MetriKon 2015 held in Cologne in November 2015, which constitute a collection of theoretical studies in the field of software measurement and case reports on the application of software metrics in companies and universities.

Schmietendorf, A. (Hrsg.):

Eine praxisorientierte Bewertung von Architekturen und Techniken für Big Data

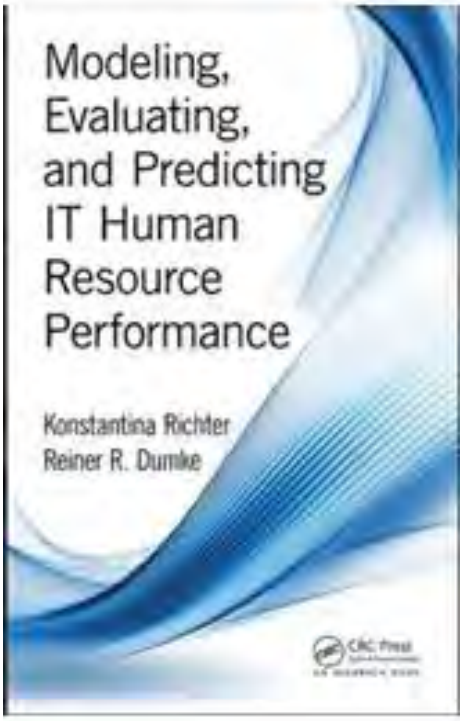
(110 Seiten) Shaker-Verlag Aachen, März 2015 ISBN 978-3-8440-2939-0

This book describes the system aspects of Big Data software infrastructures form a industrial/ practical point of view.

Konstantina Richter, Reiner Dumke:

Modeling, Evaluating and Predicting IT Human Resource Performance

CRC Press, Boca Raton, Florida, 2015 (275 pages)



This book explains why it is essential to account for the human factor when determining the various risks in the software engineering process. The book presents an IT human resources evaluation approach that is rooted in existing research and describes how to enhance current approaches through strict use of software measurement and statistical principles and criteria.

Dumke, R., Schmietendorf, A., Seufert, M., Wille, C.:

Handbuch der Softwareumfangsmessung und Aufwandschätzung

Logos Verlag, Berlin, 2014 (570 Seiten), ISBN 978-3-8325-3784-5

Eine vollständige Beschreibung der COSMIC Function Point Methode mit zahlreichen industriellen Anwendungen und Erfahrungen.

This book shows an overview about the current software size measurement and estimation approaches and methods. The essential part in this book gives a complete description of the COSMIC measurement method, their application for different systems like embedded and business software and their use for cost and effort estimation based on this modern ISO size measurement standard.

Software Measurement & Data Analysis Addressed Conferences

January 2018

CPP 2018:
7th ACM SIGPLAN International Conference on Certified Programs and Proofs
January 7 - 13, 2018, Los Angeles, USA
see: https://popl18.sigplan.org/track/CPP-2018

SWQD 2018:
Software Quality Days
January 16-19, 2018, Vienna, Austria
see: https://2018.software-quality-days.com/

SOFSEM 2018:
44th International Conference on Current Trends in Theory and Practice of Computer Science
January 29 - February 2, 2018, Krems, Austria
see: http://www.sofsem.cz/sofsem18/

February 2018

ICSEFM 2018:
20th International Conference on Software Engineering and Formal Methods
February 1 - 2, 2018, Melbourne, Australia
see: https://waset.org/conference/2018/02/melbourne/ICSEFM

ISEC 2018:
11th Innovation in Software Engineering Conference
February 9 - 11, 2018, Hyderabad, India
see: https://isoft.acm.org/isec2018/

ICEASE 2018:
20th International Conference on Empirical Assessment in Software Engineering
February 19 - 20, 2018, Paris, France
see: https://waset.org/conference/2018/02/paris/ICEASE/abstracts

ASQ 2018:
Lean and Six Sigma Conference
February 26 - 27, 2018, Phoenix, Arizona, USA
see: https://asq.org/conferences/six-sigma

March 2018

Big Data Car Data 2018:
Automobilwoche Konferenz
March 13, 2018, Munich, Germany
see: http://www.automobilwoche-konferenz.de/

REFSQ 2018:	**22th International Working Conference on Requirements Engineering: Foundation for Software Quality** March 19-22,, 2018, Utrecht, Netherlands **see: https://refsq.org/2018/welcome/**
ENASE 2018:	**13th International Conference on Evaluation of Novel Approaches to Software Engineering** March 23 - 24, 2018, Funchal, Madeira, Portugal see: http://www.enase.org/
SDDCS 2018:	**Software-Defined Data Computing and Storage workshop** March 25, 2018, Williamsburg, VA, USA see: https://sddcs.github.io/2018/sddcs2018.html
BigDataService 2018:	**IEEE Big Data Service 2018** March 26 - 29, 2018, Bamberg, Germany see: http://www.big-dataservice.net/

April 2018

Programming 2018:	**Programming 2018** April 9 - 12, 2018, Nice, France see: https://2018.programming-conference.org/
ICST 2018:	**11th International Conference on Software Testing, Verification & Validation** April 9 - 13, 2018, Västeras, Sweden see: http://www.es.mdh.se/icst2018/
BPMEA 2018:	**Business Process Management & Enterprise Architecture** April 9 - 13, 2018, Pau, France, see: http://www.cs.unibo.it/sacbpmea18/
ICPE 2018:	**9h ACM/SPEC International Conference on Performance Engineering** April 9 -13 2018, Berlin, Germnay see: https://icpe2018.spec.org/
FASE 2018:	**21th International Conference on Fundamental Approaches to Software Engineering** April 14 - 20, 2018, Thessaloniki, Greece see: https://www.etaps.org/index.php/2018/fase
ETAPS 2018:	**European Join Conference on Theory & Practice of Software** April 14 - 20, 2018, Thessaloniki, Greece see: http://www.etaps.org/index.php
CeCMG 2018:	**Enterprise Computing Conference (ECC)** April 17 - 19, 2018, Hamburg, Germany see: http://www.cecmg.de/
SOFTENG 2018:	**International Conference on Advances and Trends in Software Engineering** April 22 - 26, 2018, Athen, Greece see: http://iaria.org/conferences2018/SOFTENG18.html

iqnite 2017: *(not 2018)*
Software Quality Conference
April 24 26, 2017, Düsseldorf, Germany
see: https://www.iqnite-conferences.com/de/

CIbSE 2018:
Ibero-American Conference on Software Engineering
April 23 - 27, 2018, Bogota, Colombia
see: http://cibseconference.org/

WPDAI 2018:
The Web Conference
April 23 - 27, 2018, Lyon, France
see: https://www2018.thewebconf.org/

EMEA 2017: *(not 2018)*
PMI Global Congress 2017 - EMEA
April 29 - 30, 2017, Rome, Italy
see: http://congresses.pmi.org/emea2017

May 2018

ICSA 2018:
13th International Conference on Software Architecture
April 30 – May 4, 2018, Seatle, USA
see: http://icsa-conferences.org/2018/

STAREAST 2018:
Software Testing Analysis & Review Conference
April 29 - May 4, 2018, Orlando, FL, USA
see: http://stareast.techwell.com/

ISMA 2018:
15th ISMA Conference of the IFPUG
May 9 - 11, 2018, Roma, Italy
see: http://it-cisq.org/isma-15-conference/

ODSC East 2018:
Open Data Science Conference East
May 1 - 4, 2018, Boston, USA
see: https://www.odsc.com/boston

eMetrics 2017: *(not 2018)*
eMetrics Summit
May 15 - 18, 2017, San Francisco, USA
see: https://www.emetrics.org/sanfrancisco/2017/

MSR 2018:
Working Conference on Mining Software Repositories
May 28 - 29, 2018, Gothenburg, Sweden
see: https://2018.msrconf.org/#/home

ICPC 2018:
International Conference on Program Comprehension
May 27 - 28, 2018, Gothenburg, Sweden
see: https://conf.researchr.org/home/icpc-2018

QUEST 2018:
Quality Engineering Software and Testing Conference & EXPO
May 21 – 25, 2018, San Antonio, Texas, USA
see: http://qaiquest.org/2018/

ECASE 2017: *(not 2018)*
First International Workshop on Establishing a Community-Wide Infrastructure for Architecture-Based Software
May 20 - 28, 2017, Buenos Aires, Argentina
see: http://design.se.rit.edu/ECASE/

XP 2018:
International Conference on Agile Software Development
May 21-25, 2018, Porto, Portugal
see: https://www.agilealliance.org/xp2018/

ICGSE 2018:
13th International Conference on Global Software Engineering
May 27 - 29, 2018, Gothenburg, Sweden
see: https://conf.researchr.org/home/icgse-2018

SEAMS 2018:
13th International Symposium on Software Engineering for Adaptive and Self-Managing Systems
May 28 - 29, 2018, Gothenburg, Sweden
see: https://conf.researchr.org/track/seams-2018/seams-2018-papers

ICAMDS 2018:
International Conference on Applied Mathematics and Data Science
May 29 - 31, 2018, Hangzhou, China
see: http://www.icamds.com/

ICSE 2018:
40th International Conference on Software Engineering
May 27 - June 3, 2018, Gothenburg, Sweden
see https://www.icse2018.org/

June 2018

EJC 2018:
28th International Conference on Information Modeling and Knowledge Bases
June 4 - 8, 2018, Riga, Latvia
see: http://ejc-conference.org/index.html

ICWE 2018:
International Conference on Web Engineering
June 5 - 8, 2018, Caseres, Spain
seehttp://icwe2018.webengineering.org/

OSS 2018:
International Conference on Open Source Systems
June 8 - 10, 2018, Athen, Greece
see: https://www.oss2018.org/

SPICE 2017:
(not 2018)
16th International SPICE Conference
June 9 - 10, 2017, Dublin, Ireland
see: http://www.spiceconference.com/

ICSEA 2018:
19th International Conference on Software Engineering Advances
June 11 - 12, 2018, Copenhagen, Denmark
see:
https://waset.org/conference/2018/06/copenhagen/ICSEA

SERA 2018:
ACIS Conference on Software Engineering Research, Management and Applications
June 13 - 15, 2018, Kunmin, China
see: http://www.acisinternational.org/sera2018/

IEEE DSC 2018:	**IEEE International Conference on Data Science in Cyberspace** June 18 - 21, 2018, Guangzhou, China see: http://www.ieee-dsc.org/2018/
VDA Automotive SYS Conference 2017: *(not 2018)*	**Quality Management for Automotive Software-based Systems and Functionality** June 19 - 21, 2017, Berlin, Germany see: http://vda-qmc.de/fileadmin/redakteur/Software/sys/
Big Data 2018:	**Big Data Analysis and Data Mining** June 20 - 21, 2018, Rome, Italy see: https://conference.researchbib.com/view/event/74030
IMMM 2018:	**International Conference on Advances in Information Mining and Management** July 22 - 26, 2018, Barcelona, Spain see: http://www.iaria.org/conferences2018/CfPIMMM18.html
ICWS 2018:	**IEEE 25th International Conference on Web Services** June 25 - 30, 2018, Seatlle, USA see: http://icws.org/2018/
CLOUD 2018:	**IEEE 11th International Conference on Cloud Computing** June 25 - 30, 2018, Seattle, USA see: http://www.thecloudcomputing.org/2018/
SERVICES 2018:	**IEEE 14th World Congress on Services** June 25 - 30, 2018, Seattle, USA see: http://www.servicescongress.org/2018/
BigData 2018:	**Big Data Congress** June 25 - 30, 2018, Seattle, USA see: http://www.bigdatacongress.org/2018/
SCC 2018:	**15th International Conferences on Service Computing** June 25 - 30, 2018, Seattle, USA see: http://thescc.org/2018/index.html
AIMS 2018:	**7th International Conference on AI and Mobile Services** June 25 - 30, 2018, Seattle, USA see: http://www.ai1000.org/2018/
ICIOT 2018:	**3rd International Conference on Internet of Things** June 25 - 30, 2018, Seattle, USA see: http://www.iciot.org/2018/index.html

July 2018

ISSQ 2018:	**Ninth International Symposium on Software Quality** July 2 - 5, 2018, Melbourn, Australia see: http://sq.covenantuniversity.edu.ng/
BCD 2018:	**Big Data, Cloud Computing & Data Science Engineering** July 10 - 12, 2018, Yonago, Japan, see: http://www.acisinternational.org/bcd2018/

ISSTA 2018 : **International Symposium on Software Testing and Analysis**
July 16, 2018, Amsterdam, The Netherlands
see: https://conf.researchr.org/home/issta-2018

CSCE 2018: **15th International Conference on Software Engineering Research and Practice**
July 17 - 20, 2017, Las Vegas, Nevada, USA
see: http://americancse.org/events/csce2017/conferences/serp17

MCCSIS 2018: **International Conference on Big Data Analytics, Data Mining and Computational Intelligence**
July 17 - 20, 2018, Madrid, Spain
see: http://mccsis.org/

ICDSE 2018: **20th International Conference on Data Science and Engineering**
July 23 - 24, 2018, Istanbul, Turkey
see: https://waset.org/conference/2018/07/istanbul/ICDSE

ICSOFT 2018: **12th International Conference on Software and Data Technologies**
July 26 - 28, 2018, Porto, Portugal
see: http://www.icsoft.org/

SERP 2018: **15th International Conference on Software Engineering Research and Practice**
July 30 - August 2, 2018, Las Vegas, Nevada, USA
see: http://americancse.org/events/csce2018/conferences/serp18/

ICOMP'18: **International Conference on Internet Computing and internet of Things**
July 30 - August 2, 2018, Las Vegas, Nevada, USA
see: https://10times.com/icomp-las-vegas

EEE'18: **International Conference on e-Learning, e-Business, Enterprise Information Systems, and e-Government**
July 30 - August 2, 2018, Las Vegas, USA
see: https://americancse.org/events/csce2018/conferences/eee18/

GCC'18: **International Conference on Grid, Cloud, and Cluster Computing**
July 30 - August 2, 2018, Las Vegas, USA
see: https://americancse.org/events/csce2018/conferences/gcc18/

August 2018

ICBDM 2018: **International Conference on Big Data-Driven Management**
August 4 - 6, 2018, Beijing, China
see: http://www.icbdm.net/

icABCD'18: **International Conference on Advances in Big Data, Computing and Data Communication System**
August 6 - 7, 2018, Durban, South Africa
see: https://www.ieee.org/conferences_events/conferences/

AGILE 2018:	**Annual North American Agile Conference** August 6 - 10, 2018, San Diego, CA, USA see: https://www.agilealliance.org/agile2018/
RE 2018:	**26th IEEE International Requirement Engineering Conference** August 20 - 24, 2018, Banff, Alberta, Canada see: http://re18.org/
Euromicro DSD/ SEAA 2018:	**Software Engineering & Advanced Application Conference** August 29 - 31, 2018, Prague, Czech Republic see: http://dsd-seaa2018.fit.cvut.cz/seaa/

September 2018

ASE 2018:	**Automated Software Engineering** September 3 - 7, 2018, Montpellier, France see: http://ase2018.com/
QEST 2018:	**15th International Conference on Quantitative Evaluation of Systems** September 4 - 7, 2018, Beijing, China see: http://www.qest.org/qest2018/
EuroAsiaSPI² 2018:	**European Systems & Software Process Improvement and Innovation Conference,** September 5 - 7, 2018, Bilbao, Spain see: http://2018.eurospi.net/
ODSC Europe 2018:	**Open Data Science Conference Europe** September 19 - 22, 2018, London, UK see: https://odsc.com/london
IWSM-MENSURA 2018:	**Common International Conference on Software Measurement** September 19 - 20, 2018, Beijing, China see: http://www.iwsm-mensura.org/

October 2018

ESEIW 2018:	**Empirical Software Engineering International Week** October 8 - 12, 2018, Oulu, Finland see: https://eseiw2018.wixsite.com/oulu
ISMA 2018:	**16th ISMA Conference of the IFPUG** October 18, 2018, Sao Paulo, Brazil see: https://bfpug.wordpress.com/2018/01/19/metricas-2018-isma-16/
SAM Summit 2018:	**Conference on Software Asset Management** October 22 - 24, 2018, Chicago, USA see: http://www.ecpmedia.com/samsummit.html
ODSC 2018:	**Open Data science Conference Europe**

October 13 - 14, 2018, London, UK
see: https://www.odsc.com/london

November 2018

ESEC/FSE 2018:
European Software Engineering Conference and **Symposium on the Foundation of Software Engineering**
November 4 - 9, 2018, Lake Buena Vista, FL, USA
see: https://2018.fseconference.org/

CSEE&T 2017:
(next conf. 2019)
30th Conference on Software Engineering Education and Training
November 7 - 9, 2017, Savannah, Georgia
see: http://www.cseet2017.com/index.html

ESAPI 2018:
Workshop Evaluation of Service-APIs
November, 2018, Munich, Germany
see: https://blog.hwr-berlin.de/schmietendorf/

ASQT 2018:
Arbeitskonferenz Softwarequalität, Test und Innovation
November , 2018, Graz, Austria
see: http://www.asqt.org/

IEEE ICDM 2018:
IEEE International Conference on Data Mining
November 17 - 20 , 2018, Singapore
see: http://icdm2018.org/

PROFES 2018:
International Conference on Product Focused Software Process Improvement
November 28 - 30, 2018, Wolfsburg, Germany
see https://profes2018.wordpress.com/

December 2018

IEEM 2018:
International Conference on Industrial Engineering and Engineering Management
December 16 - 19, 2018, Bangkok, Thailand
see: http://ieem.org/public.asp?page=home.htm

see also:

- http://www.acisinternational.org/newconferences.html
- https://www.acm.org/conferences
- https://www.ieee.org/conferences_events/index.html

COMMUNITIES

GI-Fachgruppe Software-Messung und Bewertung

http://fg-metriken.gi.de/

(Measurement News Online)

Common Software Measurement International Consortium

http://cosmic-sizing.org

Deutschsprachige Anwender-gemeinschaft für Software-Metrik und Aufwandschätzung

http://www.dasma.org

International Software Bench-marking Standard Group (ISBSG)

https://www.isbsg.org

Central Europe Computer Measurement Group (ceCMG)

http://www.cecmg.de

Metrics Association's International Network (MAIN)

http://*www.mai-net.org*

Finnish Software Measurement Association (FISMA)

http://www.fisma.fi/in-english/

Netherlands Software Metrics users Association (NESMA)

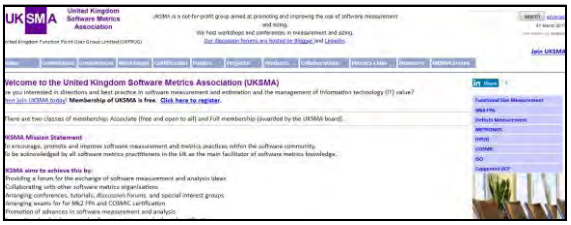
http://www.nesma.org/

Asociacion Espanola de Metricas de Software

http://www.aemes.org/

United Kongdom Software Metrics Association (UKSMA)

http://www.uksma.co.uk



Gruppo Utenti Function Point Italia - Italian Software Metrics Association (GUFPI - ISMA)

http://www.gufpi-isma.org

Anwenderkonferenz Software-qualität und Test (ASQT)

http://www.asqt.org

MEASUREMENT SERVICES

Software Measurement Laboratory (SML@b)

http://www.smlab.de

International Function Point Users Group (IFPUG)

http:www.ifpug.org

Practical Software & Systems Measurement

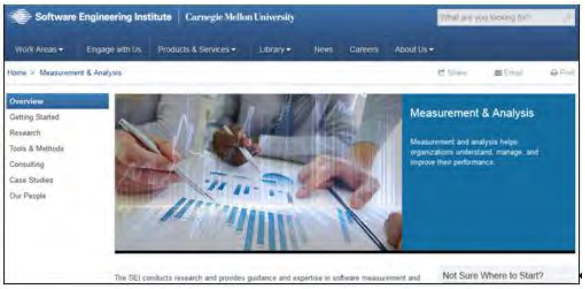
www.psmsc.com/:

Computer Measurement Group (CMG)

http://www.cmg.org

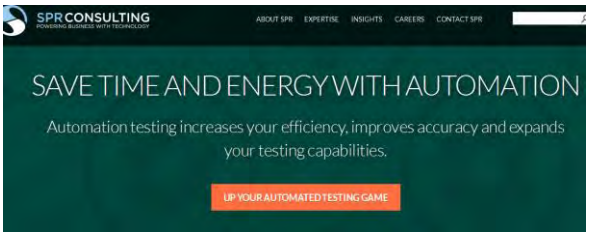
Software Engineering Institute (SEI)

www.sei.cmu.edu/measurement/



Software Productivity Research (SPR)

http://www.spr.com/



McCabe & Associates

http://www.mccabe.com

SQS Gesellschaft für Software-Qualitätssicherung

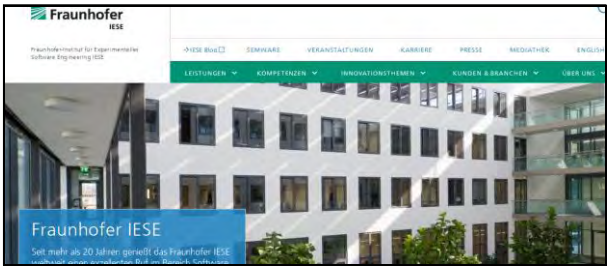
http://www.sqs.de

Quantitative Software Management (QSM)

http://www.qsm.com/

Fraunhofer Institute for Experimental Software Engineering (IESE)

https://www.iese.fraunhofer.de/



National Institute of Standards and Technology (NIST)

https://www.nist.gov/el

SOFTWARE MEASUREMENT INFORMATION

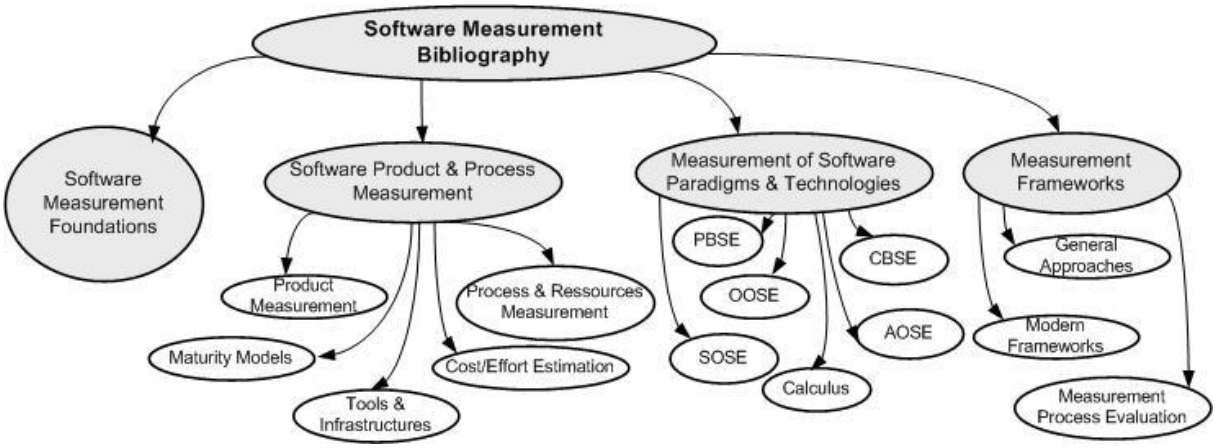
Software Measurement Bibliography

See our overview about software metrics and measurement in the Bibliography at

http://fg-metriken.gi.de/bibliografie.html

including any hundreds of books and papers

Bibliography Structure:



Software Measurement & Wikipedia

Help to qualify the software measurement knowledge and intentions in the world wide web:

Software Engineering Body of Knowledge (SWEBOK)

http://*www.swebok.orgl*

Project Management Body of Knowledge (PMBOK)

http://www.pmbook.org

SOFTWARE MEASUREMENT NEWS

VOLUME 23 2018 NUMBER 1

CONTENTS

Announcements 2

Conference Reports 7

News Papers 17

Abdalla Bala, Alain Abran
Impact Analysis of Multiple Imputation on Effort Estimation Models with the ISBSG Repository of Software Projects 17

Andreas Schmietendorf, André Nitze
Empirische Untersuchung zum Einsatz von SaaS-Lösungen bei Klein- und mittelständischen Unternehmen (KMUs) 35

New Books on Software Measurement 49

Conferences Addressing Measurement Issues 55

Metrics in the World-Wide Web 63

ISSN 1867-9196