Deggendorfer Forum zur
digitalen Datenanalyse e.V. (Hrsg.)

GoBD und Big Data

Neue Herausforderungen für die digitale Datenanalyse

ESV ERICH SCHMIDT VERLAG

ERICH
SCHMIDT
VERLAG

GoBD und Big Data

Neue Herausforderungen für die digitale Datenanalyse

Herausgegeben vom
Deggendorfer Forum zur digitalen Datenanalyse e. V.

Mit Beiträgen von
Dr. David Christen, Dominik Fischer, Anton Grening,
Dr. Markus Grottke, Holger Klindtworth, Martin Landvoigt,
Dr. Sascha Mehlhase, Dr. Lars Meyer-Pries, Günter Müller,
Wolf-Dietrich Richter, Prof. em. Dr. Hartmut J. Will

ERICH SCHMIDT VERLAG

Weitere Informationen zu diesem Titel finden Sie im Internet unter
ESV.info/978 3 503 16544 5

Gedrucktes Werk: ISBN 978 3 503 16543 8
eBook: ISBN 978 3 503 16544 5

Alle Rechte vorbehalten
© Erich Schmidt Verlag GmbH & Co. KG, Berlin 2015
www.ESV.info

Ergeben sich zwischen der Version dieses eBooks
und dem gedruckten Werk Abweichungen,
ist der Inhalt des gedruckten Werkes verbindlich.

Vorwort

Das Deggendorfer Forum 2015 stand unter dem Thema: „Datenanalyse: Objektivität, Transparenz und Vertrauen durch Big Data?!“
Die Beiträge in diesem 11. Tagungsband des Deggendorfer Forums basieren auf Vorträgen aus Wirtschaft, Forschung und Praxis. Die Ausgabe bietet ein weiteres Mal die Möglichkeit sich mit den verschiedenen Gesichtspunkten der präsentierten Inhalte auseinandersetzen zu können.

Die Eröffnungsrede des 11. Deggendorfer Forum zur digitalen Datenanalyse 2015 hielt, wie vor 10 Jahren, Prof. Dr. Hart Will aus Vancouver.
Vielen Anwendern von ACL ist er als Begründer und Entwickler der „Audit Command Language“ ein Begriff. Seinen Forschungen verdanken die Prüfer ein Werkzeug aber auch eine Methodik zur Datenanalyse, die bis heute einen Meilenstein in der digitalen Datenanalyse darstellen. Hart Will geht in seinem Eröffnungsvortrag kritisch auf Problembereiche der Datenanalyse ein, die sich seiner Meinung nach auch im Zeitalter von Big Data nicht grundlegend geändert haben.

Anschließend folgte Dr. Markus Grottke von der Universität Passau mit seinem Vortag „Rechnungslegungsbasierte Unternehmensbewertung unter Einbezug unstrukturierter Informationen – ein Markovkettenansatz“. In seinem Beitrag wird eine Lösung offeriert, um Unternehmensbewertungen mit relativ geringen Datenvoraussetzungen fundiert vorzunehmen und zeigt gleichzeitig, auf welche Weise bei der Unternehmensbewertung zudem bedeutsame qualitative Erkenntnisse in den Bewertungsprozess einbezogen werden können.

Die nächsten beiden Beiträge basieren auf Bachelorarbeiten aus der Hochschule zu Themen der Datenanalyse. In der ersten Arbeit „Journal Entry Testing - ein praxisorientierter Ansatz unter Verwendung der Netzwerkstruktur“, wurde die bereits im Deggendorfer Forum, von Prof. Mochty vorgestellte Methode zu grafentheoretische Analyse beim Journal Entry Test, von Dominik Fischer in Zusammenarbeit mit Wolf-Dietrich Richter der BDO AWT, in eine Machbarkeitsstudie umgesetzt und prototypisch implementiert.
Die Arbeit von Anton Grening und Günter Müller unter dem Thema „Datenanalyse als Vorbote für die zukunftsorientierte Risikobewertung im Mittelstand“ beschreibt die Möglichkeiten der ‚Predictive Analytics‘. In einer Machbarkeitsstudie werden den theoretischen Anforderungen und Problemen der Praxis gegenübergestellt.

Es lohnt sich gerade beim Thema Datenanalyse immer wieder, einen Blick über den Zaun zu werfen und zu sehen, wie in diesem Fall Physiker mit den Herausforderungen großer Datenmengen umgehen. Alleine die Datenmengen, die am CERN in Genf generiert und verarbeitet werden, übersteigen schon jedes Vorstellungsvermögen. Interessant und sehr aufschlussreich sind aber auch die Technologien und Methoden, die dort eingesetzt werden, um den hohen fachlichen Anforderungen der Kunden gerecht zu werden. Es ist sicher kein Zufall, dass Wirtschaft und Industrie sehr gerne Mitarbeiter aus dieser Welt für ihre Datenanalyseprojekte rekrutieren. Einen Einblick in diese Welt gibt Dr. Sascha Mehlhase, Mitarbeiter am Lehrstuhl für Elementarteilchenphysik an der LMU München, in seinem Beitrag „Die Teilchen-Nadel im Daten-Heuhaufen finden".

Aus den jüngst in Kraft getretenen GoBD (*Grundsätze zur ordnungsmäßigen Führung und Aufbewahrung von Büchern, Aufzeichnungen und Unterlagen in elektronischer Form sowie zum Datenzugriff*) ergeben sich sowohl für Prüfer als auch die Softwareindustrie neue Aufgaben, um den neuen zumindest aber neu formulierten Ansprüchen der Finanzverwaltung Rechnung zu tragen. Dr. Lars Meyer-Pries von der DATEV eG übernimmt es, diese neue Sichtweise für unser Forum aufzubereiten und zusammenzufassen.

Abschließend bezieht Holger Klindtworth in seinem Beitrag, der zu den Herausforderungen der GoBD im Zusammenhang mit Big Data aus dem Blickwinkel des Fachausschuss für Informationstechnologie des IDW, Position.

Bei allen Referenten und Mitwirkenden möchte ich mich an dieser Stelle persönlich sowie im Namen des Vereins recht herzlich für ihr großartiges Engagement bedanken ebenso auch für die Mühe mit der sie Ihr Wissen und ihre Erfahrungen in diesen Tagungsband eingebracht haben. Ohne ihre Unterstützung wäre die Herausgabe dieses Tagungsbandes nicht möglich gewesen.

Mein besonderer Dank richtet sich an die Kooperationspartner: dab: GmbH, BDO AG, DATEV eG und der Technische Hochschule Deggendorf. Für die Anpassung der schriftlichen Beiträge an ein einheitliches Layout bedanke ich mich bei Herrn Galetzka, Student der Technischen Hochschule Deggendorf, sowie beim Erich Schmidt Verlag.

Georg Herde Deggendorf, im Oktober 2015

Inhaltsverzeichnis

Vorwort **5**
Inhaltsverzeichnis **7**

Prof. Dr. Hart Will
Erhöhte Objektivität, Transparenz und Vertrauen mit Big Data?! 11
1 Abstrakt 13
2 Einleitung 14
3 Daten und Big Data 27
4 Objektivität 33
5 Transparenz 40
6 Vertrauen 46
7 Fazit 53
8 Prüfungsimplikationen 54
9 Zusammenfassung und Schluss 55
10 Literaturverzeichnis 56

Dr. David Christen, Dr. Markus Grottke, Martin Landvoigt
Rechnungslegungsbasierte Unternehmensbewertung unter Einbezug unstrukturierter Informationen – ein Markovkettenansatz 57
1 Problemstellung 59
2 Das theoretische Fundament der rechnungslegungsbezogenen Unternehmensbewertung 61
3 Ein einfaches Anwendungsbeispiel zur Ermittlung eines rechnungslegungsbezogenen Unternehmenswertes 65
4 Ein Vorschlag zum Einbezug unstrukturierter Informationen: Das Modell von Christen und Grottke (2015) 72
5 Zwei Anwendungsbeispiele zum Modell von Christen und Grottke (2015) 75
6 Ausblick 80
7 Literaturverzeichnis 81

Wolf-Dietrich Richter, Dominik Fischer
Journal Entry Testing - Ein praxisorientierter Ansatz unter Verwendung der Netzwerkstruktur 85
1 Einleitung 85
2 Theoretische Einführung 90
2.1 Die Prüfung des Buchungsjournals 85
2.2 Pflichten und Möglichkeiten durch die GDPdU und GoBD 90
2.3 Grundlagen der Graphentheorie und deren Algorithmen 91
2.4 Entstehung der Graphen-Datenbanken 92
2.5 Auszüge aus der Visualisierung und Oberflächengestaltung 92
3 Anforderungen an die Analyse-Software 94
3.1 Anforderungen an die Prüfungssoftware (Herde/Töller) 2012 94
3.2 Funktionen der netzwerkanalytischen Prüfungssoftware 94
3.3 Einsatz effizienter Usability 95
3.4 Anforderungen durch große Datenmengen 96
3.5 Integration moderner Methoden und Verfahren 97
4 Ein praxisorientiertes Softwaredesign 98
4.1 Architektur des Gesamtkonzeptes 98
4.2 Präsentation im Internet-Browser 100
4.3 Visualisierung des Netzwerks 102
5 Neue Prüfungsansätze durch Data Mining 103
6 Diskussion der Ergebnisse 105

Günter Müller, Anton Grening
Datenanalyse als Vorbote für die zukunftsorientierte Risikobewertung im Mittelstand 107
1 Bedeutung Mittelstand 109
2 Wertorientierte Unternehmensführung im Mittelstand 112
2.1 Grundsätzliches 112
2.2 Handlungsbedarf 112
2.3 Weitere Motivationspunkte 113
3 Analysebereiche 114
4 Aktuelle Entwicklungen in der Datenanalyse 115
4.1 Big Data 115
4.2 Predictive Analytics 117
4.3 GRC-Management 119
5 Analysebeispiele 120
6 Fazit 123

Dr. Sascha Mehlhase
Die Teilchen-Nadel im Daten-Heuhaufen finden 125
1 Abstrakt 127
2 Einführung 128
2.1 Experimentelle Methoden 129
2.2 Big Data 130
2.3 Datenanalyse 130
2.4 Blick auf die Wirtschaft 131
3 Weiterführende Informationen 133

Dr. Lars Meyer-Pries
GoBD und E-Bilanz im Spannungsfeld zwischen Digitalisierungs-streben und Akzeptanzanspruch 135
1 Einführung und Zusammenhang: GoBD, E-Bilanz, Datenanalyse, Big Data, Vertrauen und Akzeptanz 137
2 Status Quo: GoBD und E-Bilanz 141
2.1 GoBD 141
2.2 E-Bilanz 146
3 Hürden und Spannungsfelder aus Sicht des IT-Herstellers 149
4 Ausblick mit zukünftigen Herausforderungen 151

Holger Klindtworth
GoBD und Big Data 153
1 IT und Steuern 155
2 Die GoBD – Ausgewählte Aspekte 156
3 Zusammenspiel der GoBD mit anderen steuerlichen Regelungen 162
4 Big Data 163
5 GoBD und Big Data 165

Erhöhte Objektivität, Transparenz und Vertrauen mit Big Data?!

Prof. Em. Hartmut J. Will
Dipl.-Kfm., Ph.D, CPA, CMA

Inhaltsübersicht

1 Abstrakt
2 Einleitung
3 Daten und Big Data
4 Objektivität
5 Transparenz
6 Vertrauen
7 Fazit
8 Prüfungsimplikationen
9 Zusammenfassung und Schluss
10 Literaturverzeichnis

1 Abstrakt

„Big Data“ (ohne rationale Erklärungen) ist eine Ideologie von Informationsmonopolisten, -oligarchen und -technologen, die algorithmisch und datentechnisch fixiert, aber erkenntnistheoretisch unkritisch sind und uns weismachen wollen, dass globale Vernetzung durch sie für uns „sharing“ bedeutet. In Wahrheit geschieht sie aber zu ihren Gunsten und auf unsere Kosten. Daten sind natürlich schon immer Bausteine unseres Wissens, seit Menschen Beobachtungen machen und kritisch darüber nachdenken. Dieses Sammeln und Nachdenken jetzt programmierten Monitoren und undurchsichtigen Algorithmen zu überlassen „enthauptet“ die Geistesgeschichte der Menschheit. Hinter jeder Beobachtung (allgemein als Daten verstanden) steht eine kognitive Anforderung, Hypothese oder Vermutung, die zu formulieren man keinem noch so künstlich intelligenten Monitor oder (Such)-Mechanismus überlassen kann. Es geht viel zu oft darum, nur spekulative Hinweise, nützliche Trends oder pragmatische Argumente anstatt glaubwürdiges Wissen zu gewinnen, und sie im Eigeninteresse als objektive, transparente und vertrauensvolle Informationen zu verkaufen. Andererseits sind Daten und Informationen kommunizierbar und von kritischen und rationalen Menschen als logisch richtig oder falsch bzw. als glaubwürdig oder unglaubwürdig bewertbar. Maschinell entdeckte Korrelationen und ihr unkritisch programmierter Austausch zwischen Menschen oder Computern reichen nicht aus, weil sie nicht rational erklären, sondern sogar algorithmisch verblenden können. Folglich sind weder erhöhte Objektivität noch größere Transparenz und noch weniger verbessertes Vertrauen durch nicht aufklärende „Big Data“ zu gewinnen. Wir erliegen einer gefährlichen Illusion und einem geistes-geschichtlichen und politischen Irrtum größten Ausmaßes, wenn wir uns der Big-Data-Ideologie unkritisch ergeben und wichtige erkenntnistheoretische sowie ethische Unterscheidungen unterlassen, die wir hier am Beispiel von englisch-sprachigen Beiträgen zum Thema illustrieren und kommentieren.

Schlüsselwörter: Big Data, Daten-Governance, Informationssystem-Theorie, Prüfungstheorie

2 Einleitung

Big Data ist ein moderner *terminus technicus* mit einer scheinbar kognitiven Bedeutung, die eine ideologische Orientierung derjenigen verschleiert, welche versuchen, wohlwollend, kriminell oder „unschuldig“ Kontrolle über unser Wissen und Verhalten zu erlangen. Ohne erkenntnistheoretische und wissenschaftstheoretische Basis sind „kleine/small“ und besonders „massenhafte/big“ Daten ein Mythos. Wenn „die Wahrheit“ das darstellt, was „auf dem Bildschirm“ zu sehen ist, der von jenen kontrolliert ist, die auch „die Wolke“ kontrollieren, dann vernebeln wir unser kulturelles und philosophisches Erbe seit der Aufklärung. Alle bisher gewonnenen aufklärenden Einsichten und Erkenntnisse scheinen von Technokraten und Kontrolleuren der Informationstechnologie bewusst vergessen oder ignoriert zu sein, um ihre persönlichen, politischen, sozialen und wirtschaftlichen Interessen „big-data-ideologisch“ durchzusetzen.

> *Ideologie ist definiert als „der Komplex von Doktrin, Mythos, Glaube usw., die ein Individuum, soziale Bewegung, Institution, Klasse oder große Gruppe leitet... sowie solch ein Komplex von Doktrin, Mythos, usw. mit Bezug auf irgendeinen politischen und sozialen Plan, wie z.B. Faschismus, mitsamt den Mitteln zu ihrer Durchsetzung.“ (Google Suche auf Englisch)*

Daten sind noch keine *Informationen*, aber dokumentierte Beobachtungen, Messungen oder Vermutungen mit bestimmten (subjektiven) Intentionen, die (digital) erfasst, verschlüsselt, vernetzt und mithilfe von Meta-Daten beschrieben, erklärt und getestet werden können. Informationen sind zwar auch kein „sicheres“ *Wissen*, aber zu bestimmtem Zwecken (digital) gefilterte oder synthetisierte Daten, die zwar auch digital ermittelt, verschlüsselt, verdichtet, vernetzt und mithilfe von Meta-Informationen beschrieben, erklärt und getestet werden können, die Welt bzw. Umwelt aber nicht „automatisch“ objektivwahr, transparent oder vertrauenswürdig machen. Wissen impliziert ebenfalls keine *Gewissheit*, repräsentiert aber kritisch getestete Informationen als rationale Überzeugungen von Menschen mit eigenen Interessen, (die auch nicht immer vertrauenswürdig sind); und *Weisheit* scheint das intelligenteste Gut im Überlebenskampf in der Welt aus globaler Sicht oder innerhalb einer bestimmten Umgebung (Umwelt) aus lokaler Sicht zu sein, aber kann sie wirklich aufklärerisch - anstatt von ideologisch! – durch die Gleichung ***internet* + *internet of things* = *wisdom of the earth*** adäquat beschrieben werden? - Vergleiche dazu die folgende Umschreibung:

> *Weisheit ist die Qualität oder Zustand, weise zu sein; Wissen, was wahr und richtig ist verbunden mit richtigem Ermessen in Bezug auf Handlung; Klugheit, kritisches Urteilsvermögen oder Verständnis.*[1]

Es lohnt sich, die Situation als Erklärung-Rahmen für computer-gestützte Informationssysteme innerhalb eines Erkenntnisbereiches oder „Wirklichkeitsausschnitts" (Tetens) zu erklären und zu illustrieren. Wir können somit einen Daten-, Informations- und Wissenskontext erkenntnistheoretisch, kognitiv und technologisch unterscheiden und gleichzeitig in Beziehung zu einander zu setzen. (Vgl. Will, 2006)

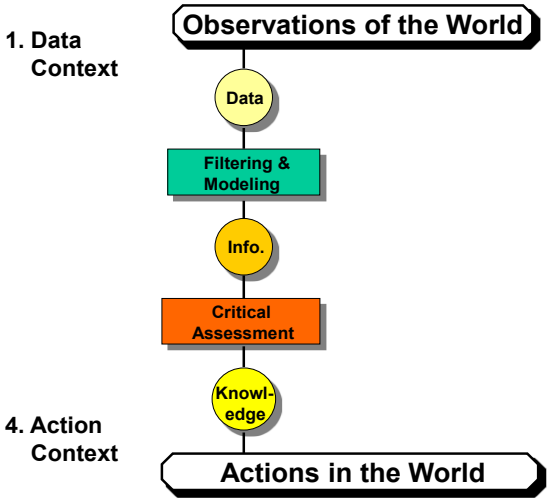

Abb. 1: Informationssystem-Rahmen

Wenn wir unser Verständnis der Welt auf digitales Kontrollieren (Monitoring), Kodieren, Verarbeiten und Darstellen von Daten und auf ihre algorithmische Manipulation beschränken, können wir im Netz unserer eigenen (digitalisierten) Mythen gefangen sein, ohne die Beziehungen zu Ethik, Privatsphäre, Realität, Sicherheit, Subjektivität, Transparenz, Vertrauen und Wahrheit zu sehen. Deshalb muss dieser Rahmen um seine Meta-Dimensionen erweitert werden, um *jenseits* der eigentlichen Beobachtungen, Daten, Informationen und unseres Wissens den jeweiligen Wirklichkeitsausschnitt überprüfbar beschreiben und erklären zu können, wie in Abb. 2 dargestellt.

[1] Wisdom is the quality or state of being wise; knowledge of what is true or right coupled with just judgment as to action; sagacity, discernment, or insight. (Google reference)

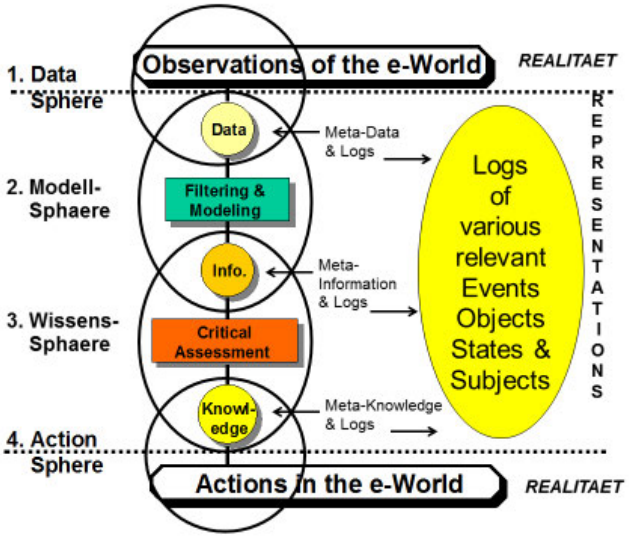

Abb. 2: Informationssystem Konzepte und Strukturen

Deswegen enthalten Logdateien Nachweise über die objektsprachliche Art der Daten-, Informations- und Wissensverarbeitung und ermöglichen *jenseits davon* metasprachliche Auswertungen und kritische Überprüfungen der jeweiligen Prozeduren und Prozesse: „Nicht in der Theorie, sondern erst in der Metatheorie erscheint die Realität." (Huebner, 1978: 71, zitiert in Tetens, 2013: 116.)

Als Wissenskomponenten können Daten also verschiedentlich analysiert, synthetisiert, verdichtet, vernetzt und „gemeinsam" (shared) benutzt werden. Die Hippie-Ideologie des „sharing everything without protection" scheint noch mitzuschwingen, wenn von *Big Data* naiv die Rede ist. Sie werden zwar auch als Wissenskomponenten in großen Mengen verstanden, sind aber bisher von einflussreichen Informationstechnokraten vornehmlich als vernetzte und zentralisierte Datenstrukturen zu den verschiedensten „gemeinnützigen" Erkenntniszwecken in den Händen von mächtigen Kontrollorganen beschrieben:

> *Die „big data" Herausforderung für Regierungsstellen und andere Institutionen in der Welt lautet: Wie können Spionage Agenturen, Militärabteilungen und Polizeiagenturen alle ihre Datenbasen in eine zentralisierte Struktur integrieren, so dass die richtigen Punkte miteinander verbunden*