



## [ Prof. Dr. Klemens Böhm // Systeme der Informationsverwaltung ]

60 Klemens Böhm ist seit 2004 Inhaber der Professur für Datenbanken und Informationssysteme am KIT. Davor war er gut zwei Jahre Professor für Angewandte Informatik/Data and Knowledge Engineering an der Otto-von-Guericke-Universität Magdeburg. Von 1998 bis 2002 war er Oberassistent an der ETH Zürich in der Datenbankgruppe, von 1993 bis 1998 wissenschaftlicher Mitarbeiter der GMD – Forschungszentrum Informationstechnik GmbH am Darmstädter Institut für Integrierte Publikations- und Informationssysteme (IPSI). Klemens Böhm promovierte 1997 an der Technischen Hochschule Darmstadt mit einer Arbeit über die Verwaltung semi-strukturierter Daten mit Datenbanksystemen. Er studierte Informatik mit Nebenfach Betriebswirtschaftslehre in Frankfurt, Darmstadt und Lissabon (Diplom von der TH Darmstadt 1993).

Derzeitige Forschungsthemen der Gruppe sind die Entwicklung von Analysetechniken für große Datenbestände und die Verwaltung wissenschaftlicher Daten. Unsere inhaltlichen Schwerpunkte sind dabei insbesondere Datenqualität, die sinnvolle Berücksichtigung von Domänenwissen und Anwenderwissen oder die Generierung synthetischer Daten. Der Lehrstuhl arbeitet viel mit anderen Lehrstühlen, seien sie Informatik, seien sie aus anderen Disziplinen, und mit Anwendern zusammen, seien sie Wissenschaftler, seien sie aus der Industrie. Klemens Böhm ist Sprecher des 2015 bewilligten DFG-Graduiertenkollegs „Energiezustandsdaten – Informatik-Methoden zur Erfassung, Analyse und Nutzung“, in dem KIT-Forscherinnen und -Forscher aus unterschiedlichen Disziplinen datengestützt an der Entwicklung effizienter, nachhaltiger, robuster und benutzerfreundlicher Energiesysteme arbeiten.

### // Überblick und Allgemeines

Die derzeitigen Forschungsthemen der Professur sind die Entwicklung von Analysetechniken für große Datenbestände und die Verwaltung wissenschaftlicher Daten. Ein wichtiges Forschungsziel ist das Erkennen von Auffälligkeiten, insbesondere in Datenbeständen hoher Dimensionalität (also beispielsweise Datenobjekte mit sehr vielen Attributen) oder in Zeitreihen bzw. Datenströmen. Daraus ergeben sich vielfältige Möglichkeiten für Anwender, beispielsweise bessere Verfahren für die Wartung von Maschinen und Anlagen. Außerdem interessiert uns die Effizienz und Skalierbarkeit der Analysetechniken. Ein konkretes Problem ist beispielsweise das Auffinden von Paaren von Zeitreihen, deren zeitliche Verläufe maximal ähnlich zueinander sind, gegeben eine sehr große Menge solcher Zeitreihen.

Während ein vorherrschendes Forschungsziel darin besteht, Auffälligkeiten möglichst gut zu erkennen, ist auch die „umgekehrte“ Frage wichtig, wie gut die zugrunde liegenden Daten sein sollten, um das Erkennen von Auffälligkeiten mit einer bestimmten Qualität zu ermöglichen. Denn die Ausgestaltung/Dimensionierung der Infrastruktur für die Erfassung der Daten hängt erheblich davon ab. Zum anderen lassen sich deutliche Verbesserungen bei jener Erkennung dadurch erzielen, dass man Domänenwis-

sen bzw. Anwenderwissen berücksichtigt – die weitgehend offene Frage ist aber, wie das in den Erkennungsprozess einfließen sollte. Schließlich geht es uns auch darum, ausgehend von bestehenden Datenbeständen Auffälligkeiten synthetisch zu generieren, um dadurch zusammen mit Anwendern ein tieferes Verständnis für sie zu bekommen.

Unsere Forschungsthemen passen gut zur thematischen Ausrichtung in der Lehre, insbesondere zu Datenbanktechnologie und zur Datenanalyse, seien es Vorlesungen, Seminare oder Praktika.

## // Ergebnisse und Erfolge

Das DFG-Graduiertenkolleg „Energiezustandsdaten – Informatik-Methoden zur Erfassung, Analyse und Nutzung“, dessen Sprecher Herr Professor Böhm ist, hat nach einer wissenschaftlichen Begutachtung mit anschließender Verlängerung um weitere 4,5 Jahre seine inhaltliche Arbeit fortgesetzt. Dazu gehört u. a. auch die Durchführung eines Workshops zum Thema ‚Energiedaten‘ als international sichtbare wissenschaftliche Veranstaltung.

Herr Dr. Edouard Fouché, ehemaliger Doktorand bei Professor Böhm, wurde von der Helmholtz-Gemeinschaft im Fachbereich „Information“ mit dem Doktorandenpreis 2020 ausgezeichnet. Der Titel seiner Doktorarbeit lautet „Estimating Dependency, Monitoring and Knowledge Discovery in High-Dimensional Data Streams“. Herr Dr. Fouché wurde außerdem ins Eliteprogramm für Postdoktorandinnen und Postdoktoranden der Baden-Württemberg-Stiftung aufgenommen. Sein Projekt „Sequential Decision Making Algorithms for Knowledge Discovery from Data Streams“ wird von der Stiftung drei Jahre lang gefördert.

## // Ausgewählte Veröffentlichungen

Florian Kalinke, Pawel Bielski, Snigdha Singh, Edouard Fouché, Klemens Böhm. An Evaluation of NILM Approaches on Industrial Energy-Consumption Data. The 12th ACM International Conference on Future Energy Systems (e-Energy 2021), (online) Turin, Italien, Juni/Juli 2021.

Vadim Arzamasov, Klemens Böhm. REDS: Rule Extraction for Discovering Scenarios. SIGMOD International Conference on Management of Data, (online) Xi’an, China, Juni 2021.

Gabriela Suntaxi, Aboubakr Achraf El Ghazi, Klemens Böhm: Preserving Secrecy in Mobile Social Networks. ACM Trans. Cyber Phys. Syst. 5(1): 5:1–5:29 (2021)

Georg Steinbuss, Klemens Böhm: Generating Artificial Outliers in the Absence of Genuine Ones – A Survey. ACM Trans. Knowl. Discov. Data 15(2): 30:1–30:37 (2021)

Georg Steinbuss, Klemens Böhm: Benchmarking Unsupervised Outlier Detection with Realistic Synthetic Data. ACM Trans. Knowl. Discov. Data 15(4): 65:1–65:20 (2021)

Fabian Laforet, Christian Olms, Rudolf Biczok, Klemens Böhm: An Ensemble Technique for Better Decisions Based on Data Streams and its Application to Data Privacy. IEEE Trans. Knowl. Data Eng. 33(12): 3662–3674 (2021)

## // Mitarbeiterinnen und Mitarbeiter

### Verwaltungspersonal

Barbara Breitenstein  
Bettina Wagner

### Wissenschaftliches Personal

Vadim Arzamasov  
Jakob Bach  
Daniel Betsche  
Pawel Bielski  
Béla Böhnke  
Adrian Englhard  
Edouard Fouché  
Marco Heyden  
Florian Kalinke  
Federico Matteucci  
Jutta Mülle  
Elaheh Ordoni  
Moritz Renftle

### Technisches Personal

Christian Möck  
Herma Teune

## // Website

[dbis.ipd.kit.edu](https://dbis.ipd.kit.edu)