



## Liebe Leserinnen, liebe Leser,

wir freuen uns, Ihnen nach einer längeren, Corona-bedingten Pause eine neue Ausgabe des Forschungsmagazins ‚Blick in die Wissenschaft‘ in der Ausgabe 44/45 präsentieren zu können.

Die Corona-Pandemie hat auch die Universität Regensburg und alle ihre Mitglieder vor große Herausforderungen gestellt. Dennoch konnten zentrale Zukunftsprojekte weitergeführt und umgesetzt werden. So stellt vor allem die Gründung unserer neuen Fakultät für Informatik und Data Science (FIDS) einen wahren Meilenstein in der Geschichte und Entwicklung der Universität Regensburg dar. Als größtes Strukturprojekt seit der Gründung der Fakultät für Medizin vor 30 Jahren ist unsere Informatikfakultät ein Zukunftsprojekt von weitreichenden Dimensionen. Mit der neuen strategischen Schwerpunktsetzung im Bereich Informatik und Data Science und vor allem auch der Querschnittsorientierung der neuen Fakultät sieht sich die Universität Regensburg sehr gut gerüstet, ihre bisherigen Stärken in diesen Bereichen zu bündeln, weiter auszubauen und zu

vertiefen. Schließlich sind *Digital Transformations* als eines der vier Gestaltungsfelder und Zukunftsthemen in unserem *Hochschulentwicklungsplan 2025* fest verankert. Dieses Gestaltungsfeld adressiert die neue Fakultät ebenso wie den Bereich *Integrated Sciences in Life, Health, and Disease* als ein weiteres Schwerpunktgebiet unserer Universität.

Die Grundsatzbeschlüsse in den Gremien der Universität Regensburg im Sommer und Herbst 2019 zur Einrichtung der neuen Fakultät erfolgten nach einer vorhergehenden Phase intensiver Planungen dann letztlich fast zeitgleich mit der Regierungserklärung des Bayerischen Ministerpräsidenten Dr. Markus Söder am 10. Oktober 2019 und der Verkündung der Hightech Agenda Bayern. Unterstützt und beschleunigt durch die Mittel der Hightech Agenda Bayern konnte der Auf- und Ausbau der Fakultät für Informatik und Data Science zügiger umgesetzt werden, nachdem die neue Fakultät im März 2020 formal gegründet und im Laufe des WS 2021/22 aus sich heraus handlungs- und

funktionsfähig wurde. Im Mai 2022 konnten wir gemeinsam mit Ministerpräsident Dr. Markus Söder und Staatsminister für Wissenschaft und Kunst Markus Blume den offiziellen Kickoff für die Fakultät begehen. Dass dieser komplexe Prozess im Kontext der Herausforderungen der Corona-Pandemie vollzogen und abgeschlossen werden konnte, ist ein Zeichen für die Bedeutung dieser gesamtuniversitär-strategischen Maßnahme und für den Rückhalt für das Großprojekt in der universitären Gemeinschaft.

Im Laufe des Gründungsprozesses ist es gelungen, die verschiedenen Informatiknahen und -interessierten Kräfte der Universität an einen Tisch zu bringen und gemeinsam ein zukunftsorientiertes Konzept für die Fakultät zu entwickeln. Ein externes Gutachten mit hochrangiger Expertise skizzierte und evaluierte 2019 wesentliche inhaltliche Schwerpunkte und Strukturierungen für die neue Fakultät, an denen sich in den Jahren 2019-2021 die von Vizepräsident Prof. Dr. Nikolaus Korber geleitete Gründungskommission in der

konkreten Arbeit zum Aufbau der Fakultät orientierte. In insgesamt 15 Berufungsverfahren wurden die ersten neuen Professuren in der Fakultät zügig besetzt – ein Prozess, der in Kürze abgeschlossen sein wird. Im Besetzungsprozess hat sich vor allem auch gezeigt, wie attraktiv die Neugründung einer Fakultät und die Möglichkeiten zur Mitgestaltung und zum Aufbau neuer Strukturen für Wissenschaftlerinnen und Wissenschaftler sind und wie viel Zukunftspotential von unserer neuen Fakultät ausgeht. So konnten wir zum Wintersemester 2023/24 130 Studierende für den B.Sc. Informatik und den B.Sc. Data Science begrüßen.

Im vorliegenden Heft von ‚Blick in die Wissenschaft‘ möchten wir Ihnen nunmehr vor allem die Forschungsaktivitäten der Fakultät für Informatik und Data Science näher vorstellen. Dabei beglückwünsche ich die Fakultät, dass sie bereits eineinhalb Jahre nach ihrer vollständigen Handlungs- und Funktionsfähigkeit und während der weiteren Planungen zum Aufbau und der Ausarbeitung ihrer Studiengänge insbe-

sondere im Master-Bereich ein so vielfältiges Themenheft zu ihren aktuellen Forschungsarbeiten vorlegen konnte.

Das facettenreiche und vielfältige Themenspektrum dieses Sonderhefts illustriert, wie die Fakultät für Informatik und Data Science die an der Universität Regensburg bisher vorhandenen IT-Kompetenzen erfolgreich bündelt und in die Zukunft gerichtet erweitert. Sie ermöglicht die essentielle interdisziplinäre Vernetzung mit der gesamten Universität, von den Geistes- und Sozialwissenschaften bis zu den Natur- und Lebenswissenschaften. Die Beiträge verdeutlichen, wie interdisziplinäre Forschung das Fundament starker methodischer und fachlicher Grundlagen weiterentwickelt und wie die bisherigen Informatik-Schwerpunkte der Universität Regensburg (Computational Science, Informationswissenschaft, Medieninformatik, Wirtschaftsinformatik) erfolgreich in die neue Fakultät überführt werden konnten und die Querschnittsorientierung unterstützen.

Das vorliegende Heft mit seinem Schwerpunkt auf aktuellen Forschungsar-

beiten begleitet den im Wintersemester 2023/24 erfolgten Start der beiden grundständigen Bachelor-Studiengänge Informatik und Data Science. Ein kurzer Überblicksbeitrag zur Lehre in der neuen Fakultät zeigt anschaulich die bereits gewachsene Vielfalt der Informatikstudiengänge und die intensive und gelebte Verbindung von Forschung und Lehre auch an dieser neuen Fakultät.

Unsere neue Fakultät leistet hervorragende Arbeit und ich bin sicher, dass Ihnen die nachfolgenden Seiten einen spannenden Einblick in die verschiedenen Facetten der FIDS geben werden. Ein ganz besonderes Dankeschön möchte ich an dieser Stelle an das gesamte Dekanat der Fakultät für Informatik und Data Science und insbesondere an Forschungsdekanin Prof. in Dr. Meike Klettke richten, die für diese Sonderausgabe die Koordinationsarbeit der vorliegenden Ausgabe federführend übernommen hat.

Prof. Dr. Udo Hebel

Präsident der Universität Regensburg

**Blick in die Wissenschaft  
Forschungsmagazin  
der Universität Regensburg**

ISSN 0942-928-X

Heft 44/45

31. Jahrgang

**Herausgeber**

Prof. Dr. Udo Hebel

Präsident der Universität Regensburg

**Redaktionsleitung für diese Ausgabe**

Prof.in Dr. Meike Klettke / Fakultät für Informatik und Data Science

**Redaktionsbeirat**

Prof. Dr. jur. Christoph Althammer

Prof. Dr. rer. nat. Ferdinand Evers

Prof. Dr. rer. nat. Stefan Friedl

Prof. Dr. rer. nat. Mark W. Greenlee

Prof. Dr. theol. Andreas Merkt

Prof. Dr. phil. Omar W. Nasim

Prof. Dr. rer. nat. Klaus Richter

Prof. Dr. rer. pol. Daniel Rösch

Prof. Dr. med. Ernst Tamm

Prof. Dr. paed. Oliver Tepner

Prof. Dr. phil. Christiane Heibach

Universität Regensburg

93040 Regensburg

Telefon +49 941 9432300

Telefax +49 941 9433310

**Verlag**

Universitätsverlag Regensburg GmbH

Leibnizstraße 13, 93055 Regensburg

Telefon +49 941 78785-0

Telefax +49 941 78785-16

info@univerlag-regensburg.de

www.univerlag-regensburg.de

Geschäftsführer: Dr. Albrecht Weiland,

Felix Weiland M.A.

**Abonnement-service**

bestellung@univerlag-regensburg.de

**Anzeigenleitung**

Larissa Nevecny

MME-Marquardt

info@mme-marquardt.de

**Herstellung**

Universitätsverlag Regensburg GmbH

info@univerlag-regensburg.de

**Einzelpreis € 7,00**

**Doppelheft € 14,00**

**Jahresabonnement**

bei zwei Ausgaben pro Jahr

**€ 10,00 / ermäßigt € 9,00**

Für Schüler, Studierende und Akademiker/innen im Vorbereitungsdienst (inkl. 7% MwSt.) zzgl. Versandkostenpauschale

€ 1,64 je Ausgabe. Bestellung beim Verlag.

Für **Mitglieder des Vereins der Ehemaligen Studierenden der Universität Regensburg e.V.**, des **Vereins der Freunde der Universität Regensburg e.V.** und des **Vereins ehemaliger Zahnmedizinstudenten Regensburg e.V.** ist der Bezug des Forschungsmagazins im Mitgliedsbeitrag enthalten.

Für **Mitglieder des Vereins der Ehemaligen Studierenden der Universität Regensburg e.V.**, des **Vereins der Freunde der Universität Regensburg e.V.** und des **Vereins ehemaliger Zahnmedizinstudenten Regensburg e.V.** ist der Bezug des Forschungsmagazins im Mitgliedsbeitrag enthalten.

Für **Mitglieder des Vereins der Ehemaligen Studierenden der Universität Regensburg e.V.**, des **Vereins der Freunde der Universität Regensburg e.V.** und des **Vereins ehemaliger Zahnmedizinstudenten Regensburg e.V.** ist der Bezug des Forschungsmagazins im Mitgliedsbeitrag enthalten.

Für **Mitglieder des Vereins der Ehemaligen Studierenden der Universität Regensburg e.V.**, des **Vereins der Freunde der Universität Regensburg e.V.** und des **Vereins ehemaliger Zahnmedizinstudenten Regensburg e.V.** ist der Bezug des Forschungsmagazins im Mitgliedsbeitrag enthalten.

Für **Mitglieder des Vereins der Ehemaligen Studierenden der Universität Regensburg e.V.**, des **Vereins der Freunde der Universität Regensburg e.V.** und des **Vereins ehemaliger Zahnmedizinstudenten Regensburg e.V.** ist der Bezug des Forschungsmagazins im Mitgliedsbeitrag enthalten.

Für **Mitglieder des Vereins der Ehemaligen Studierenden der Universität Regensburg e.V.**, des **Vereins der Freunde der Universität Regensburg e.V.** und des **Vereins ehemaliger Zahnmedizinstudenten Regensburg e.V.** ist der Bezug des Forschungsmagazins im Mitgliedsbeitrag enthalten.

Für **Mitglieder des Vereins der Ehemaligen Studierenden der Universität Regensburg e.V.**, des **Vereins der Freunde der Universität Regensburg e.V.** und des **Vereins ehemaliger Zahnmedizinstudenten Regensburg e.V.** ist der Bezug des Forschungsmagazins im Mitgliedsbeitrag enthalten.

Für **Mitglieder des Vereins der Ehemaligen Studierenden der Universität Regensburg e.V.**, des **Vereins der Freunde der Universität Regensburg e.V.** und des **Vereins ehemaliger Zahnmedizinstudenten Regensburg e.V.** ist der Bezug des Forschungsmagazins im Mitgliedsbeitrag enthalten.

Für **Mitglieder des Vereins der Ehemaligen Studierenden der Universität Regensburg e.V.**, des **Vereins der Freunde der Universität Regensburg e.V.** und des **Vereins ehemaliger Zahnmedizinstudenten Regensburg e.V.** ist der Bezug des Forschungsmagazins im Mitgliedsbeitrag enthalten.

Für **Mitglieder des Vereins der Ehemaligen Studierenden der Universität Regensburg e.V.**, des **Vereins der Freunde der Universität Regensburg e.V.** und des **Vereins ehemaliger Zahnmedizinstudenten Regensburg e.V.** ist der Bezug des Forschungsmagazins im Mitgliedsbeitrag enthalten.

Für **Mitglieder des Vereins der Ehemaligen Studierenden der Universität Regensburg e.V.**, des **Vereins der Freunde der Universität Regensburg e.V.** und des **Vereins ehemaliger Zahnmedizinstudenten Regensburg e.V.** ist der Bezug des Forschungsmagazins im Mitgliedsbeitrag enthalten.

Für **Mitglieder des Vereins der Ehemaligen Studierenden der Universität Regensburg e.V.**, des **Vereins der Freunde der Universität Regensburg e.V.** und des **Vereins ehemaliger Zahnmedizinstudenten Regensburg e.V.** ist der Bezug des Forschungsmagazins im Mitgliedsbeitrag enthalten.

Für **Mitglieder des Vereins der Ehemaligen Studierenden der Universität Regensburg e.V.**, des **Vereins der Freunde der Universität Regensburg e.V.** und des **Vereins ehemaliger Zahnmedizinstudenten Regensburg e.V.** ist der Bezug des Forschungsmagazins im Mitgliedsbeitrag enthalten.

Für **Mitglieder des Vereins der Ehemaligen Studierenden der Universität Regensburg e.V.**, des **Vereins der Freunde der Universität Regensburg e.V.** und des **Vereins ehemaliger Zahnmedizinstudenten Regensburg e.V.** ist der Bezug des Forschungsmagazins im Mitgliedsbeitrag enthalten.

Für **Mitglieder des Vereins der Ehemaligen Studierenden der Universität Regensburg e.V.**, des **Vereins der Freunde der Universität Regensburg e.V.** und des **Vereins ehemaliger Zahnmedizinstudenten Regensburg e.V.** ist der Bezug des Forschungsmagazins im Mitgliedsbeitrag enthalten.

Für **Mitglieder des Vereins der Ehemaligen Studierenden der Universität Regensburg e.V.**, des **Vereins der Freunde der Universität Regensburg e.V.** und des **Vereins ehemaliger Zahnmedizinstudenten Regensburg e.V.** ist der Bezug des Forschungsmagazins im Mitgliedsbeitrag enthalten.

Für **Mitglieder des Vereins der Ehemaligen Studierenden der Universität Regensburg e.V.**, des **Vereins der Freunde der Universität Regensburg e.V.** und des **Vereins ehemaliger Zahnmedizinstudenten Regensburg e.V.** ist der Bezug des Forschungsmagazins im Mitgliedsbeitrag enthalten.

Für **Mitglieder des Vereins der Ehemaligen Studierenden der Universität Regensburg e.V.**, des **Vereins der Freunde der Universität Regensburg e.V.** und des **Vereins ehemaliger Zahnmedizinstudenten Regensburg e.V.** ist der Bezug des Forschungsmagazins im Mitgliedsbeitrag enthalten.

Für **Mitglieder des Vereins der Ehemaligen Studierenden der Universität Regensburg e.V.**, des **Vereins der Freunde der Universität Regensburg e.V.** und des **Vereins ehemaliger Zahnmedizinstudenten Regensburg e.V.** ist der Bezug des Forschungsmagazins im Mitgliedsbeitrag enthalten.

Für **Mitglieder des Vereins der Ehemaligen Studierenden der Universität Regensburg e.V.**, des **Vereins der Freunde der Universität Regensburg e.V.** und des **Vereins ehemaliger Zahnmedizinstudenten Regensburg e.V.** ist der Bezug des Forschungsmagazins im Mitgliedsbeitrag enthalten.

Für **Mitglieder des Vereins der Ehemaligen Studierenden der Universität Regensburg e.V.**, des **Vereins der Freunde der Universität Regensburg e.V.** und des **Vereins ehemaliger Zahnmedizinstudenten Regensburg e.V.** ist der Bezug des Forschungsmagazins im Mitgliedsbeitrag enthalten.

Für **Mitglieder des Vereins der Ehemaligen Studierenden der Universität Regensburg e.V.**, des **Vereins der Freunde der Universität Regensburg e.V.** und des **Vereins ehemaliger Zahnmedizinstudenten Regensburg e.V.** ist der Bezug des Forschungsmagazins im Mitgliedsbeitrag enthalten.

Für **Mitglieder des Vereins der Ehemaligen Studierenden der Universität Regensburg e.V.**, des **Vereins der Freunde der Universität Regensburg e.V.** und des **Vereins ehemaliger Zahnmedizinstudenten Regensburg e.V.** ist der Bezug des Forschungsmagazins im Mitgliedsbeitrag enthalten.

Für **Mitglieder des Vereins der Ehemaligen Studierenden der Universität Regensburg e.V.**, des **Vereins der Freunde der Universität Regensburg e.V.** und des **Vereins ehemaliger Zahnmedizinstudenten Regensburg e.V.** ist der Bezug des Forschungsmagazins im Mitgliedsbeitrag enthalten.

Für **Mitglieder des Vereins der Ehemaligen Studierenden der Universität Regensburg e.V.**, des **Vereins der Freunde der Universität Regensburg e.V.** und des **Vereins ehemaliger Zahnmedizinstudenten Regensburg e.V.** ist der Bezug des Forschungsmagazins im Mitgliedsbeitrag enthalten.

Für **Mitglieder des Vereins der Ehemaligen Studierenden der Universität Regensburg e.V.**, des **Vereins der Freunde der Universität Regensburg e.V.** und des **Vereins ehemaliger Zahnmedizinstudenten Regensburg e.V.** ist der Bezug des Forschungsmagazins im Mitgliedsbeitrag enthalten.

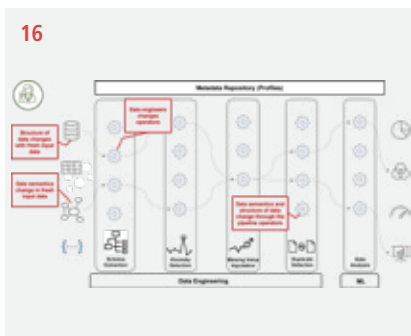
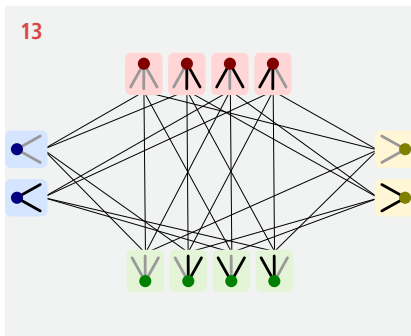
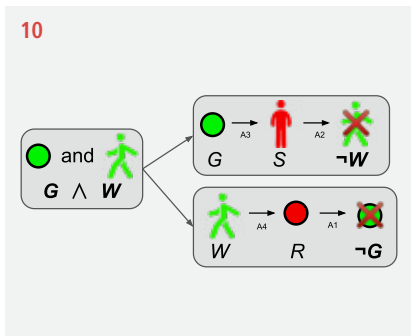
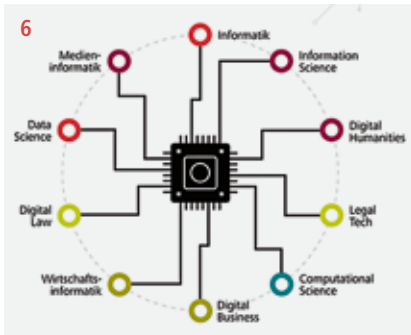


Rohstoffe  
Transporte  
Produktion

g CO<sub>2</sub>e  
492  
Pro Produkt

CO<sub>2</sub>-Emissionen  
ausgeglich

# Inhalt



**Einleitung** 5  
*Florian Erhard, Bernd Heinrich, Meike Klettke, Christian Wolff*

**Lehre an der Fakultät für Informatik und Data Science** 6  
*Florian Erhard, Udo Kruschwitz, Bernd Heinrich, Christian Wolff*

**Automatisches Beweisen: Methoden und Anwendungen** 10  
*Julie Cailler, Philipp Rümmer*

**Algorithmen und Komplexitätstheorie** 13  
*Radu Curticapean*

**Evolution in Datenbanken und Data Engineering Workflows** 16  
*Meike Klettke*

**IoT-basiertes Prozessmanagement – Mobile Benutzerführung in der digitalen Fabrik** 19  
*Stefan Schönig*

**Cyber Threat Intelligence: Gemeinschaftliche IT-Sicherheit durch den Austausch von Informationen** 23  
*Johannes Grill, Daniel Schlette, Günther Pernul*

**Kann man den Entscheidungen Künstlicher Intelligenz trauen? Zu den Auswirkungen unsicherer Daten auf die Entscheidungen Neuronaler Netze** 26  
*Thomas Krapf, Bernd Heinrich*

**Mensch vs. Maschine: Wettbewerb und Kooperation mit künstlicher Intelligenz in digitalen Märkten** 30  
*Andreas Schauer, Daniel Schnurr*

**Notfallpläne für den Ernstfall testen** 34  
*Maria Leitner*

**Maschinelles Lernen mit Anwendungen in den Naturwissenschaften** 37  
*Merle Behr, Markus Schmitt*

Automatisierte, KI-basierte Analyse von Bilddaten:

Der Lehrstuhl für Bildverarbeitung

Dorit Merhoff

40

Die Genome des Menschen – Forschungsschwerpunkte der Arbeitsgruppe für Algorithmische Bioinformatik

Birte Kehr

43

Algorithmen zum Entschlüsseln der Genregulation

Francisca Rojas Ringeling, Stefan Canzar

46

Mit Hilfe von Daten Immunprozesse entschlüsseln:

Der Lehrstuhl Computational Immunology

Florian Erhard

49

Maschinelles Lernen enthüllt den verborgenen Prozess der Tumorentstehung

Linda Hu, Andreas Lösch, Rainer Spang

52

Allgegenwärtige Mensch-Maschine-Interaktion: Entwicklung, Forschung und Infrastruktur der Medieninformatik

Raphael Wimmer, Johanna Bogon, Niels Henze, Christian Wolff

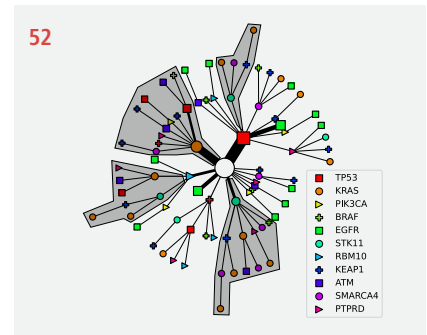
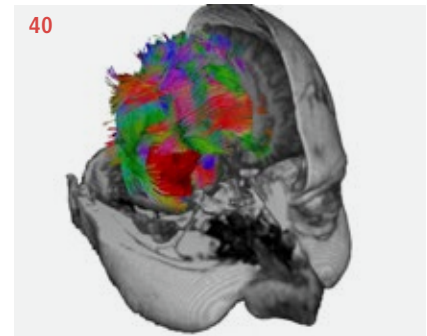
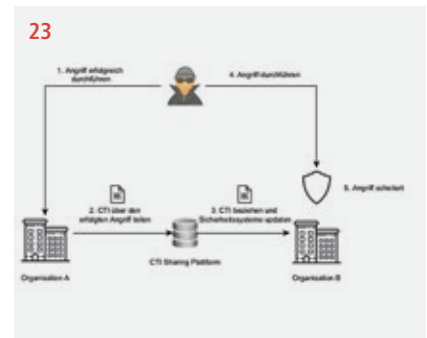
54

Wissen aus dem Internet – Genug, genau, geprüft und geeignet?

Informationswissenschaft in Regensburg

Udo Kruschwitz, Bernd Ludwig, David Elsweiler

60



## SOFTWAREENTWICKLER (M/W/D) GESUCHT

Voll- oder Teilzeit (30 -40 Std.) | Alle Fachbereiche

Hallo, **wir sind mb Support.**

Seit 20 Jahren unterstützen wir die Versicherungswirtschaft mit unseren Lösungen, u.a. Konzerte und Kunstwerke weltweit zu versichern. Als inhabergeführtes **Familienunternehmen** arbeiten wir auf Augenhöhe in einer familiären Atmosphäre. Als stark wachsendes und profitables Unternehmen bieten wir gleichzeitig **attraktive Benefits**.

**Wer bist du?** Jedes unserer rund 60 Teammitglieder bringt seine eigene Persönlichkeit, Geschichte und Perspektive mit. Für uns zählt, was dich interessiert, was dich antreibt, wie du bist. Du entscheidest, wie du bei uns mitgestalten möchtest. **Wir freuen uns auf dich.**

Die gesamte Stellenausschreibung findest du auf [www.mbsupport.de/karriere](http://www.mbsupport.de/karriere).

[www.mbsupport.de](http://www.mbsupport.de) | Friedenstraße 18 | 93053 Regensburg | +49 941 942 60 0 | mb Support GmbH

# Evolution in Datenbanken und Data Engineering Workflows

Prof. Dr. Meike Klettke

Wenn Sie den Begriff Evolution hören, denken Sie sicher an Biologie. Wir alle haben in der Schule Darwins Evolutionstheorie kennengelernt, die unter anderem beinhaltet, dass Veränderungen ständig passieren und eine Quelle der weiteren Entwicklung sind.

Den Begriff der Evolution verwenden wir auch in der Informatik. Auch hier sind Änderungen ständig auf der Tagesordnung. Programme und Daten, die nicht aktualisiert werden, altern. Das passiert, obwohl digitale Objekte wie Software und elektronisch gespeicherte Daten nicht physisch verschleißen; sie werden dennoch über die Zeit immer weniger verwendbar, wenn sie nicht kontinuierlich aktualisiert werden. Am deutlichsten wird es sicher sichtbar, wenn man einmal eine Webseite oder ein Computerprogramm findet, die seit mehreren Jahren nicht verändert wurden, sie erscheinen uns heute hoffnungslos veraltet.

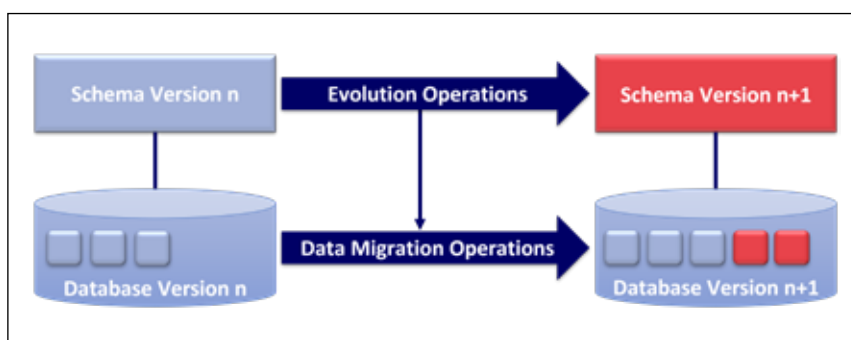
Sowohl Programme als auch Daten und deren Speicherung müssen sich also ständig weiterentwickeln, das nennen wir in der Informatik Evolution von Software und Daten. Im Mittelpunkt der Forschungsthemen des Lehrstuhls Data Engineering steht die kontinuierliche Evolution der Datenbestände. Diese ist notwendig, damit die

Daten aus der Vergangenheit, die wir auch als legacy Daten oder Altdatenbestände bezeichnen, heute noch auswertbar bleiben. Oft sind die Daten sogar wesentlich länger im Einsatz als die Software, die diese verwendet. In vielen Bereichen müssen die Daten über lange Zeit aufbewahrt werden, wie zum Beispiel bei der gesetzlich vorgeschriebenen Aufbewahrungspflicht für Rechnungsdaten. Hier ist es gefordert, dass diese Daten mindestens 10 Jahre verfügbar sein müssen. In anderen Anwendungen werden Daten noch über deutlich längere Zeiträume verwendet, auch weil deren Erstellung mit Kosten verbunden war. Vor allem aber erhofft man sich durch die Auswertung der Daten einen Erkenntnisgewinn, der sich auch daraus ergeben kann, dass man in Firmen und in der Wissenschaft Auswertungen über längere Zeiträume durchführen kann. In ökonomischen oder gesellschaftlichen Fragen möchten wir bestimmte Daten zusammen mit den Vergleichswerten des Vorjahres auswerten können; in wissenschaftlichen Experimenten, in denen wir zum Beispiel Umweltdaten erfassen, möchten wir diese über sehr lange Vergleichszeiträume auswerten können.

Teilweise sind gespeicherte Firmendaten wie auch Mess- und Experimentdaten viele Jahre alt. In einigen Forschungsinstituten reicht die Erhebung sogar etliche Jahrzehnte zurück. Nun kann man sich natürlich leicht vorstellen, dass sich seitdem einiges in der Informatik geändert hat, nicht nur die Hardware, sondern auch die Speicherformate, die Datenmodelle und die Schemata der Daten haben sich weiterentwickelt.

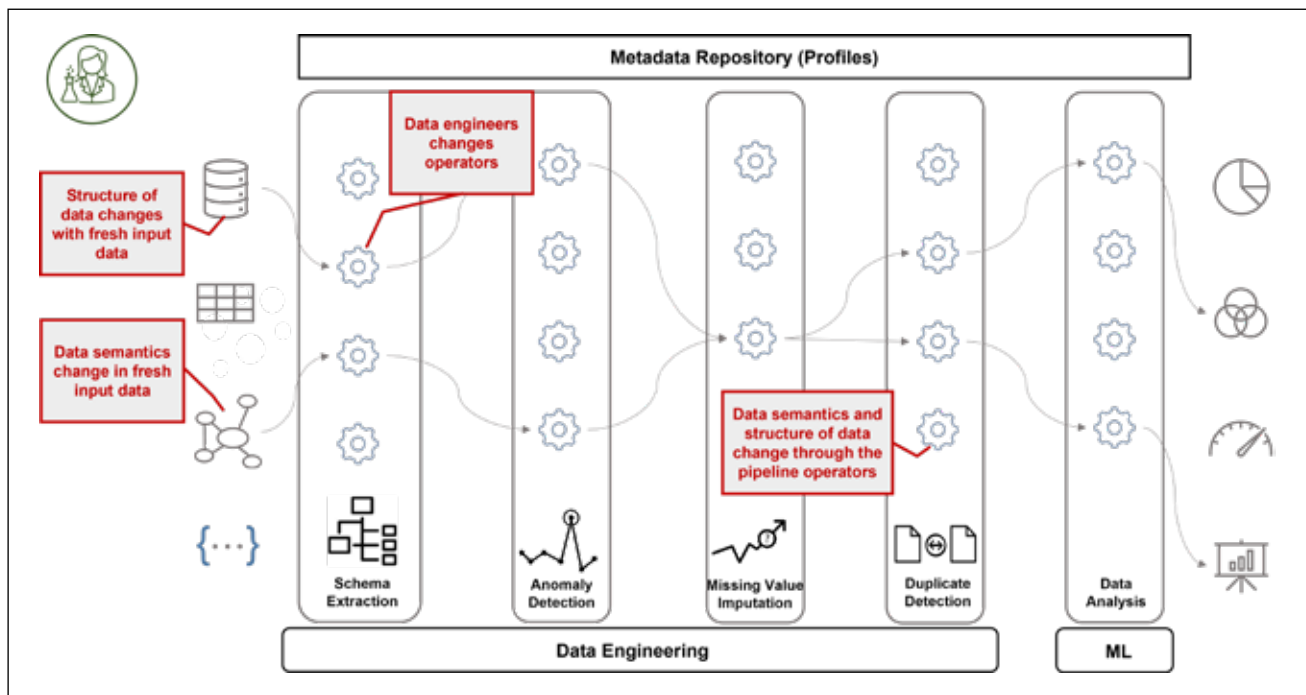
Um solche legacy Daten zu aktualisieren, entwickelt unsere Gruppe Verfahren, die Daten zwischen verschiedenen Datenmodellen transformieren und die Evolution von Daten in die aktuellen Schemata durchführen. Schematisch ist das in Abbildung 1 dargestellt. Dafür haben wir eine Evolutionssprache entwickelt, die sowohl einfache Erweiterungen und Änderungen mit den Operationen add, delete und rename durchführen kann, als auch komplexe Umstrukturierungen von Daten durch die Operationen copy, move, split und merge. Mit diesen Operationen ist ein komplettes Refactoring von Datenbanken möglich.

Die Umsetzung dieser Evolutionssprache für verschiedenen Datenbankmanagementsysteme ist eine weitere Forschungsaufgabe. So wurde die Umsetzung der Evolutionsoperationen für JSON-Daten entwickelt. JSON steht für JavaScript Object Notation und ist ein gut lesbares Datenmodell zur Darstellung von Objekten, damit ist es also sehr flexibel einsetzbar. JSON Datenbanksysteme sind Datenbankmanagementsysteme, die solche JSON Objekte performant speichern können. Das bekannteste System dieser Klasse ist MongoDB [2]. Da das Datenmodell komplexer ist als in den relationalen Datenbanken, die alle Daten tabellarisch speichern



1 Schemaevolution und Datenmigration für NoSQL-Datenbanken und Graphdatenbanken





2 Evolution von Data Engineering Workflows

und SQL als Anfragesprache dazu einsetzen, gehören JSON Datenbanken zu den sogenannten NoSQL Datenbanken. NoSQL steht dabei für »not only SQL«. Diese Datenbanksysteme sind gegenwärtig häufig im Einsatz, insbesondere Data Science Anwendungen verwenden das Speicherformat. Für die darin gespeicherten Daten entwickeln wir Verfahren zur Evolution, die eingesetzt werden können, um die Daten in neue Versionen zu migrieren und so für die Auswertung bereitzustellen.

Darüber hinaus werden bei uns am Lehrstuhl Evolutionsoperationen für Graphdatenbanken entwickelt [3]; hier erfolgt die Umsetzung für Neo4J [4]. Graphdatenbanksysteme sind eine andere Technologie zur Datenspeicherung. Ihre Bedeutung hat in den letzten Jahren stark zugenommen, Graphen werden als die natürliche Form der Datenspeicherung in der Logistik, zur Darstellung von Rechnernetzen und anderen Kommunikationswegen sowie in sozialen Netzwerken angesehen. Da immer mehr Daten als Graphen gespeichert werden, ist auch hier die Evolution, also die Überführung in die aktuellen Strukturierungen erforderlich. Wenn man solche Daten aktualisiert, dann soll das Verfahren dazu so weit wie möglich automatisiert werden. Wir arbeiten gegenwärtig an Methoden, damit die einmal für einen Datenbestand beschriebenen Evolutionsoperationen, also eine Folge aus add, rename, copy, merge

usw., auf alle Daten in einem Schritt angewendet werden können.

Das ist in Abbildung 1 schematisch für beide Klassen von Datenbanksystemen zu sehen. Entwickler:innen sollen mit dem geringstmöglichen Aufwand Daten kontinuierlich weiterentwickeln können, indem sie die gewünschten Änderungen mit der formalen Sprache (Schemaevolution) beschreiben. Der sich anschließende Schritt der Datenmigration erfolgt dann automatisch durch das Verfahren.

Ebenso ist eine ressourcenschonende Evolution möglich, die nur die benötigten Daten »on demand« aktualisiert. Das heißt also, dass erst in dem Moment, in dem Daten benötigt werden, geprüft wird, ob zur Beantwortung einer aktuellen Anfrage oder Lösen einer Aufgabe auch legacy Datenbestände existieren. Dabei wird nur der Anteil der Daten ausgewählt und aktualisiert, der für die Beantwortung der Frage benötigt wird. Es ist leicht vorstellbar, dass das technologisch etwas aufwändiger ist als das vollständige Aktualisieren des Gesamtdatenbestandes. Die dazu von uns entwickelte Methode ist ein Query Rewriting [5], bei dem eine Anfrage auf viele verschiedene (frühere) Versionen der Datenspeicherung verteilt wird und die Teilergebnisse zusammengesetzt werden. Beide Verfahren (vollständige Evolution und Evolution on demand) wurden in zwei DFG-Projekten in enger Zusammenarbeit mit den Arbeitsgruppen von Prof. Uta Störl

(FernUniversität in Hagen) und Prof. Stefanie Scherzinger (Universität Passau) für Daten in verschiedenen Datenmodellen entwickelt.

Die Weiterentwicklung der Daten ist nur ein Aspekt, warum in Informatiksystemen eine Aktualisierung notwendig sein kann. In anderen Aufgaben am Lehrstuhl entwickeln wir Verfahren für sogenannte Data Engineering Workflows. Das sind Folgen von Programmen, die Rohdaten, wie zum Beispiel vorhandene Datenbestände, Messdaten, Daten wissenschaftlicher Experimente oder empirische Befragungen aufbereiten. Dabei werden diese Daten kontrolliert, korrigiert, vervollständigt, transformiert, in einheitliche Formate integriert und dabei validiert, sodass sie auswertbar werden. Abbildung 2 zeigt schematisch, wie aus den verschiedenen verfügbaren Data Engineering Algorithmen komplette Workflows zusammengestellt werden können.

Die Auswertungen können auf verschiedene Weise erfolgen, möglich sind Datenbankabfragen, OLAP-Auswertungen, (OLAP steht für On-Line Analytical Processing und beschreibt Analysen auf großen Datenbeständen), Report-Generierungen und Analyseverfahren auf Basis von Machine Learning Methoden. Ein wichtiges Ziel ist es dabei, dass alle durchgeführten Schema- und Datenänderungen nachvollziehbar, reproduzierbar und replizierbar sind. Hierfür kommen in den aktuellen

Arbeiten am Lehrstuhl sogenannte Provenance-Techniken zum Einsatz, die sich sowohl auf die Daten als auch den Workflow selbst beziehen. Mit diesen kann für Anfrageergebnisse bestimmt werden, aus welchen Daten und durch welche Verarbeitungsschritte im Workflow diese entstanden sind. Damit werden die Ergebnisse eines Data Engineering Workflows erklärbar, ebenso werden die Ergebnisse der Workflow Evolution nachvollziehbar.

Unsere momentan laufenden Forschungsarbeiten im Bereich Data Engineering Workflows umfassen mehrere Themen: i) die Entwicklung einer Monitoringkomponente soll Änderungen in den Daten sichtbar machen und das Erkennen von Datenfehlern ermöglichen, ii) im Fall einer Evolution der Daten oder der Workflows wollen wir bestimmen können, inwieweit die Ergebnisse der Auswertung noch mit den Ergebnissen früherer Auswertungen vergleichbar sind, iii) Data Engineering Workflows können auch in der Cloud durchgeführt werden, hierbei ist die Opti-

mierung der Data Engineering Workflows für die jeweiligen Cloud Provider ein aktuelles Thema und iv) nicht zuletzt wenden wir die Data Engineering Techniken auch für spezielle Datenarten wie Eye-Tracking-Daten an.

Auch Data Engineering Workflows benötigen Verfahren zur Evolution, um im Fall von Änderungen Data Engineering Workflows zu aktualisieren und so die Workflows über lange Zeiträume und unter veränderten Rahmenbedingungen einsatzbereit zu halten. In Abbildung 2 sind die verschiedenen Arten von Änderungen, die dabei zu berücksichtigen sind, farblich hervorgehoben. Für die großen Themenbereiche der Optimierung und Evolution von Datenbeständen, sowie der Erstellung, Verwendung, Kontrolle, Protokollierung und Evolution von Data Engineering Workflows entwickeln wir in unseren gegenwärtigen Forschungsarbeiten sowohl die formalen Methoden, die zur Anwendung kommen, als auch die Tools für deren Verwendung [6].

## DREI FRAGEN ...

### *Welches Fachbuch lesen Sie gerade?*

Momentan lese ich mehrere Grundlagenbücher über die Programmiersprache Python – immer auf der Suche nach guten Erklärungen und interessanten Beispielen für die Vorlesung Programmierung I für das 1. Fachsemester im Bachelor Informatik und im Bachelor Data Science.

### *Die Lösung von welchem fachlichen Thema haben Sie sich leichter vorgestellt?*

Die Speicherung und Auswertung heterogener Daten. Datenbanksysteme können sehr gut und effizient mit tabellarischen Daten umgehen, viele Informationen sind aber entweder komplexer aufgebaut oder unregelmäßig strukturiert oder beides. Die vollständige Zuverlässigkeit von Datenbanksystemen dennoch zu garantieren, ist eine große fachliche Herausforderung.

### *Was möchten Sie in den nächsten Jahren entwickeln?*

Meine Wunschvorstellung ist es, Werkzeuge für das Data Engineering zu bauen, mit denen Anwendungsexpert:innen ihre Daten selbst bereitstellen können. Bisher erfordern diese Verfahren noch einen großen manuellen Aufwand und speziell für eine Aufgabe programmierte Skripte. Das Ziel und gleichzeitig auch der aktuelle Forschungsgegenstand unserer Gruppe ist es, Werkzeuge zu entwickeln, mit denen diese Schritte leichter ausführbar und intuitiv verständlich werden.

## Literatur

- [1] Uta Störl, Meike Klettke: Darwin: A Data Platform for NoSQLSchema Evolution Management and Data Migration, DataPlat@EDBT, 2022
- [2] MongoDB, <https://www.mongodb.com/>
- [3] Dominique Hausler, Meike Klettke and Uta Störl: A Language for Graph Database Evolution and its Implementation in Neo4j, 42nd International Conference on Conceptual Modeling (ER), Forum, 2023
- [4] Neo4j Graph Database & Analytic, Graph Database Management System, <https://neo4j.com/>
- [5] Mark Lukas Möller, Meike Klettke, Andrea Hillenbrand, Uta Störl: Query Rewriting for Continuously Evolving NoSQL Databases. Article, International Conference on Conceptual Modeling (ER conference), 2019
- [6] Meike Klettke, Uta Störl: Four Generations in Data Engineering for Data Science - The Past, Presence and Future of a Field of Science, Datenbank-Spektrum, 2022



Foto © rivat

**Prof. Dr.-Ing. habil. Meike Klettke** ist Professorin für Data Engineering an der Universität Regensburg. Sie studierte Informatik an der Universität Rostock, promovierte dort 1997 zum Thema „Akquisition von Integritätsbedingungen in Datenbanken“ und habilitierte 2007 mit ihren Arbeiten zu „Modellierung, Bewertung und Evolution von XML-Dokumentkolektionen“. Seit April 2022 leitet sie die Data Engineering Gruppe an der Fakultät für Informatik und Data Science.