

# Big Data Improves our Understanding of Diseases

What causes a disease? Why do some patients respond to a particular drug while others do not? What would the optimum therapy be for each individual patient? At TUM's Chair of Medical Informatics, Prof. Klaus A. Kuhn and his team are working to answer these questions with the help of big data.

Link

[www.imse.med.tum.de](http://www.imse.med.tum.de)

Birgit Fenzel

## Mit Big Data Krankheiten besser verstehen

Viele der häufigen und gefährlichen Krankheiten wie Krebs, Herzinfarkt, aber auch Schlaganfall und andere neurologische Erkrankungen werden durch zahlreiche Faktoren beeinflusst. Entsprechend schwierig sind Ursachenforschung, Diagnose und Therapie. Big Data könnte bei der Lösung des Problems eine große Rolle spielen. Davon ist der Medizininformatiker Prof. Klaus A. Kuhn von der TUM überzeugt. „Wenn wir über eine breite Basis von heterogenen, nach Art und Herkunft vielfältigen Daten verfügen, besteht eine gute Chance, Krankheitsverläufe, Krankheitsprävention, Diagnostik und Therapie besser zu verstehen“, betont er. Während der Behandlungen werden immer mehr technische

Geräte angewendet, die Daten produzieren. Untersuchungen wie Röntgen, CT oder MRT erzeugen eine große Menge an heterogenen Daten. Zusammen mit ärztlichen Berichten, den persönlichen Patientendaten sowie den Analyseergebnissen aus den neuen Laborverfahren, die sich als sogenannte Omics-Technologien nach der Entschlüsselung des menschlichen Genoms rasant entwickelt haben, ist das Datenvolumen seit einigen Jahren explosionsartig angewachsen.

Am Institut für Medizinische Statistik und Epidemiologie arbeitet Kuhn mit seinem Team an Konzepten und Lösungen, diese umfangreichen und heterogenen Daten aus Klinik, Forschung und Labor zu integrieren und zugänglich zu machen. Die Idee ist es, mithilfe von Informationstechnologie und analytischen Verfahren auch Vorhersagemodellen den Weg zur personalisierten Medizin zu bereiten. □

**P**rof. Kuhn, do you hope medicine can benefit from big data and data mining technologies, where algorithms and statistics are used to analyze huge data volumes?

One major benefit would be new research findings enabling medical professionals to tailor treatment precisely to individual patients. Gaining a better understanding of the way diseases develop and progress is also a key aim. To find biomarkers to predict the course of disease is also important.

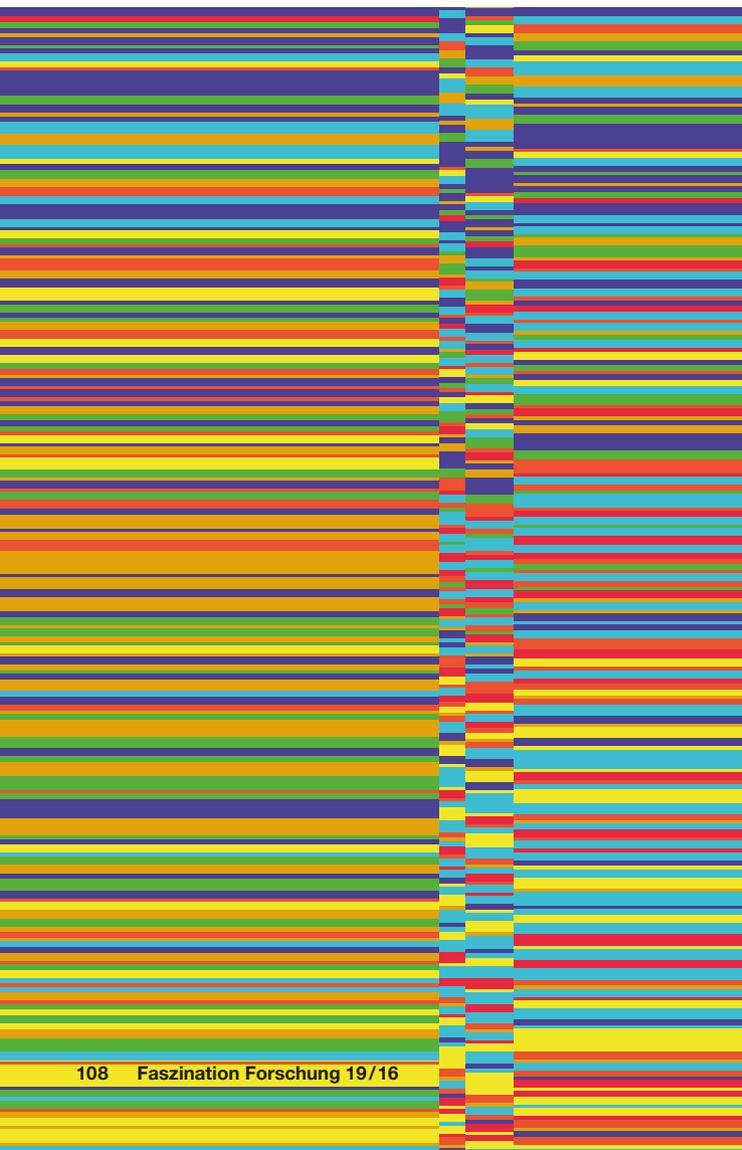
**Are there specific neurological conditions that are particularly well suited to these big data analyses, and why?**

In many cases, diseases have no single cause – there are a number of contributing factors. In neurology, this would apply to strokes, for instance, and also to multiple sclerosis, where a number of risk factors have been identified. The interplay between these various factors and their influence on the disease is complex and, to a large extent, not yet understood. But if we have access to large collections of heterogeneous data – comprising different types of data from different sources – we have a good chance of gaining a better understanding of disease progression, prevention, diagnosis

and treatment. With information technology and subsequent analyses, including predictive modeling, big data can help to pave the way for personalized medicine.

**At your institute, you are developing innovative IT concepts and solutions for translational medicine – a bench-to-bedside approach that links the lab to clinical practice and bundles knowledge, resources, expertise and techniques across disciplines. How do you get a handle on the information chaos that arises when you capture all kinds of data from the most varied of sources?**

We use various integration methods. The data warehouse concept is an important example – a warehouse in this context being a large database. It duplicates data from the various systems we work with – in IT we call this replication. This replicated data can then be queried and reorganized without touching the original sources. This type of database gives data scientists and analysts a collective view of all factors that might be relevant based on our knowledge to date. There are also other methods that do not involve data replication.



## Big Data in Medical Research

Big data describes the integration and analysis of large data volumes of high variety from different sources. An important aim is to detect patterns that are not evident in smaller or homogeneous data sets. In medicine, large data volumes are generated by “omics” technologies – that is, genomics, proteomics and metabolomics, or the study of genes, proteins and metabolites, and by imaging procedures. Sensor data from wearable devices are playing an increasingly important role for “big data”.

Neurologists and neuroscientists at TUM are exploring the potential of big data as an approach for predicting the course of disease and therapeutic outcomes. Patients respond very differently to medical treatments and physicians can lose valuable time trying to find the most effective therapy for each individual patient. As a step towards improved prediction, a database can play an important role: the basic idea is to integrate medical reports, descriptions of clinical findings, therapies, data on disease progression with imaging data (MRI, CT, etc.) and genetic information for large numbers of patients. Researchers see huge potential in this approach to gain new insights and to enable physicians to select the therapy most likely to be effective for a new patient based on the specifics (clinical findings, imaging data, genetics, etc.) of their individual case.

*“One major benefit would be new research findings enabling medical professionals to tailor treatment precisely to individual patients.”*

*Klaus A. Kuhn*

#### **How does your database work?**

Well, the aim is to mine the relevant information from these large data volumes and detect relationships and patterns by analytical methods. Data documenting the course of a disease and the possible side effects of a therapy in a specific age group may be combined with information about other factors such as information on genetics. One of the ultimate goals is to support decisions with data and knowledge at the point of care. There are successful examples, such as in the prediction of drug effects based on genetic testing. The issue then was why some patients respond to a particular drug while others do not.

#### **Your analytics and computational models call for a database that is as large and wide-ranging as possible. Why does quantity not mean quality in this case?**

We face challenges with harmonization and standardization in this area. More and more technical devices are being used in medical care that produce high volumes of data. In particular, imaging techniques such as X-ray, CT and MRI scans generate large amounts of heterogeneous data. Together with medical reports, personal patient information and analysis results from new lab techniques – omics technologies have evolved at high speed since sequencing the human genome – the last few years have seen an explosion in data volumes. In this area, as in today’s imaging procedures, standardization is a core challenge, since different devices and imaging techniques can lead to different findings, for instance.

#### **So what are the next steps, in your view?**

The main focus will be on data integration both within and across institutions, spanning high variety and heterogeneity. This requires us to harmonize and standardize both data and processes. Ethical, legal and social questions we face in today’s medical research play a key role here, especially in dealing with sensitive patient information. Data protection concepts thus are increasingly important – bringing us back to technical challenges in which we are very well positioned.

*Interview by Birgit Fenzel*




---

**Prof. Klaus A. Kuhn**

---

Klaus Kuhn holds degrees in computer science, mathematics and medicine (Dipl. Inform., Dipl. Math., Dr. med., Dr. med. habil.) from the Universities of Stuttgart, Freiburg, Tübingen and Ulm. He worked at the University Hospitals of Heidelberg and Ulm. From 1996 to 2004 he held the Chair of the Department of Medical Informatics at the University of Marburg (Philipps-Universität Marburg). In parallel to this he was Chief Information Officer at the University Hospital of Marburg. At TUM, he is a full Professor of Medical Informatics. Between 2007 and 2011 he was President and Vice-President of the German Association for Medical Informatics, Biometry and Epidemiology. Kuhn was elected a Fellow of the American College of Medical Informatics (AMIA) in 2008.