

# Skalierbare Infrastrukturen für Big Data Anwendungen



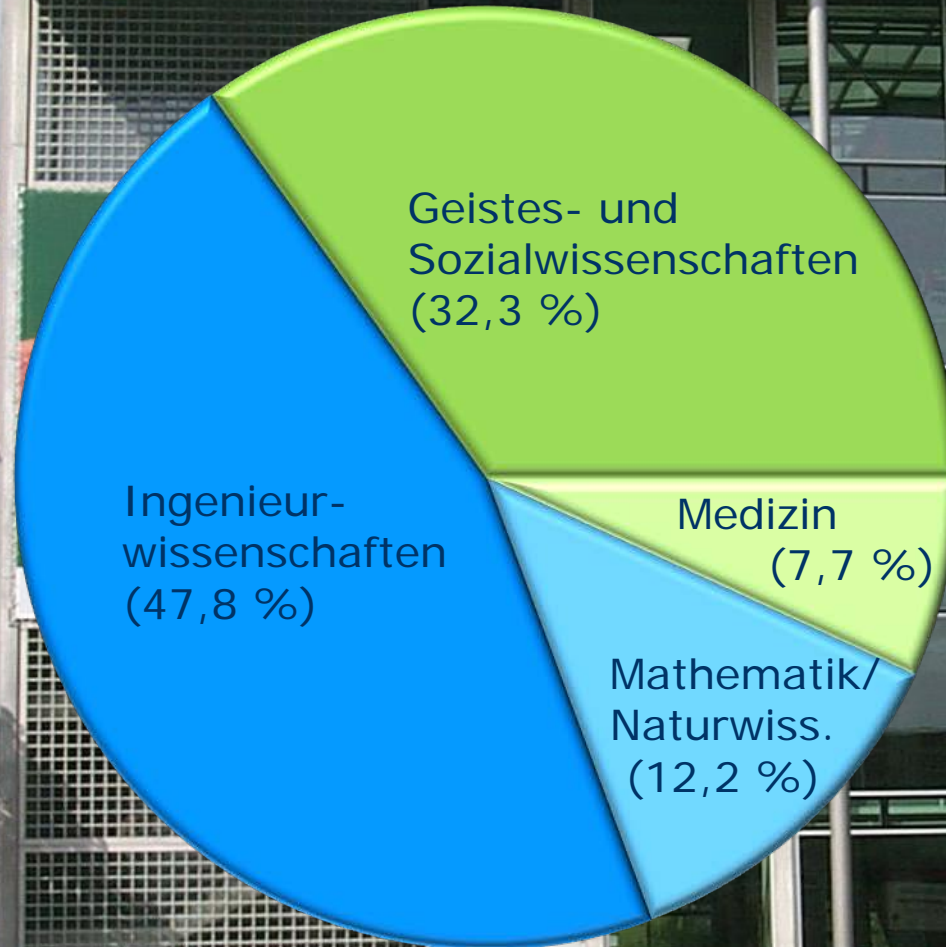




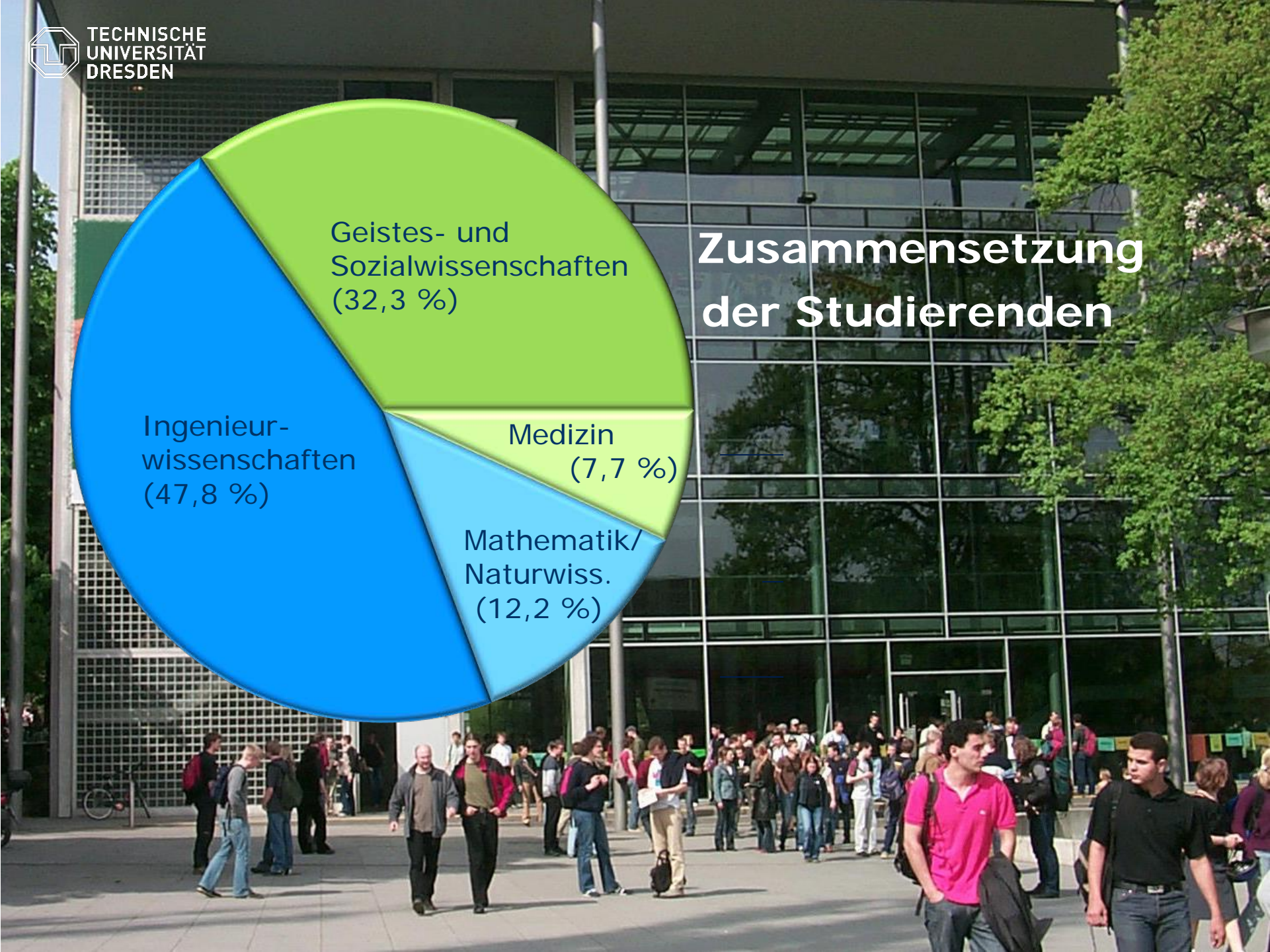
# TU Dresden: University of Excellence







## Zusammensetzung der Studierenden



# Big Data – reale Herausforderung in Wissenschaft und Wirtschaft

---

## Technologischer Fortschritt in Datenerfassung und IT-gestützter Analyse

Web 2.0

Internet of Things

Industry 4.0

Computing (HPC, Grids, Clouds)



**Groß-skalige Experimente und Simulationen** produzieren Big Data

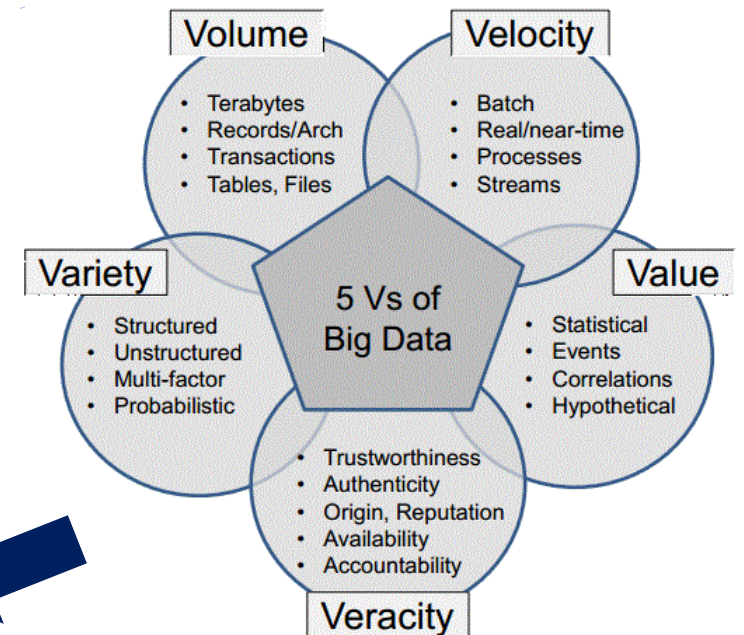
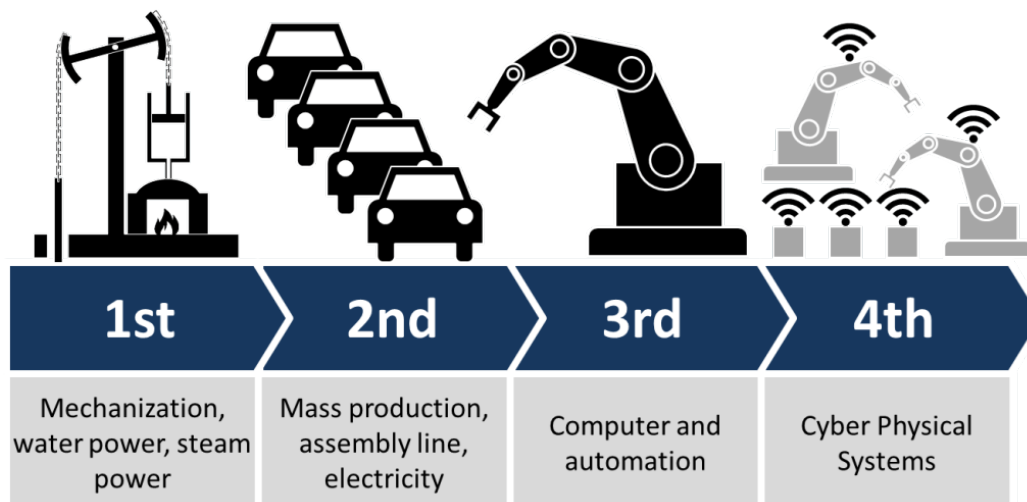
**Vereinfachter Zugang zu großen Rechner- und Speichersystemen**

**Datenflut ist Realität  
geworden!**



# Generelle Herausforderungen

- Wie integriere ich Daten aus unterschiedlichen Systemen (Standards?)
- Verständnis der zugrunde liegenden Datenbasis notwendig!
- Welche Algorithmen lassen sich überhaupt auf vorliegende Daten anwenden?
- Welche technischen Methoden lassen sich konkret nutzen?



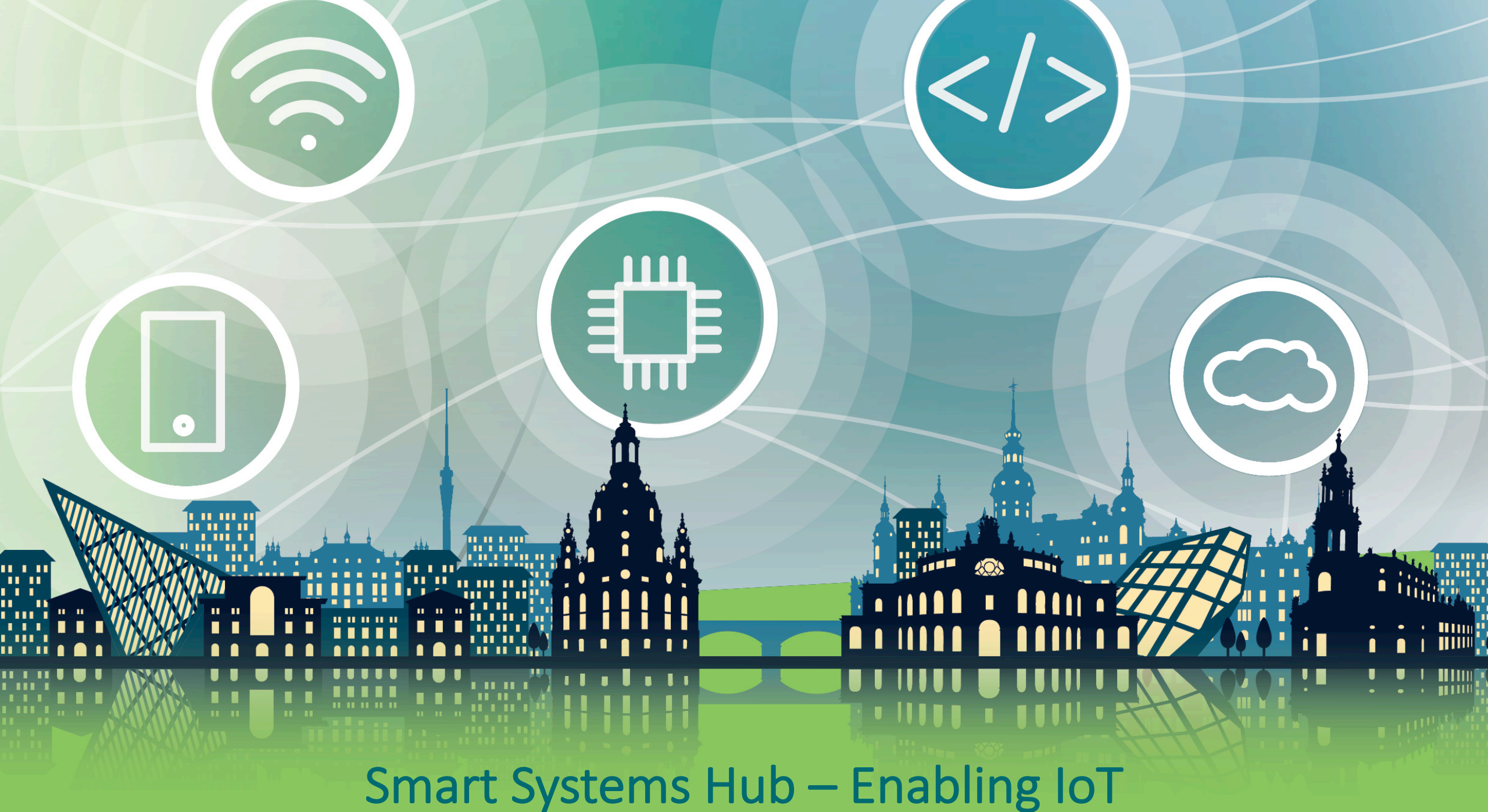
# Aktueller Stand

---

- Datensätze sind im allgemeinen **groß, hochdimensional** und **heterogen**
- Zum Erkenntnisgewinn eignen sich beispielsweise Methoden des **Maschinellen Lernens** (z.B. Cluster Analysis) oder der **Optimierung** (z.B. Local Searches), die oft aufwendig angepasst werden müssen, entweder:
  - Anpassung über die Daten (Transformationen)
  - Anpassung der Darstellungen/Visualisierungen
- Noch **keine allg. Standardverfahren** für allg. Daten verfügbar (Ansätze wie Sampling, Single Scanning, Feature Selection / Reduction, Co-Clustering, Domänentransformationen etc. vorhanden)
- **Sehr dynamisches Forschungsumfeld (Data Analytics)**

# Where is data coming from? – Mobility and Transportation

- Floating Car Data (FCD)
- Mobile Networks
- RFID data, freight transport
- Social Media
- Ticketing Systems
- Navigation Systems
- Traffic Counting, Car/Bike-Sharing
- Weather/Environmental Data
- ...



Smart Systems Hub – Enabling IoT



# Digitalisierung der Industrie und der Produkte

## Grundkomponenten eines modernen Fahrzeugs



Hardware

Sensorsysteme  
Aktoren  
Datenvorverarbeitung



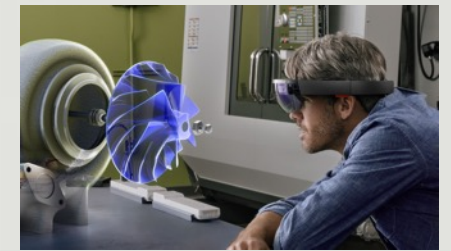
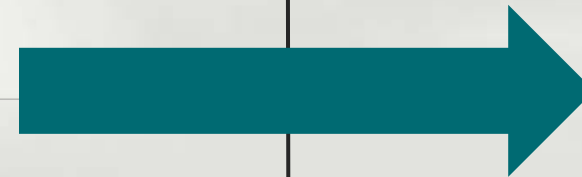
Software

Embedded Software  
High-Level Software



Connectivity

Wireless  
Low-Latency  
High-Throughput  
High Reliability



# Auf dem Weg zu IoT-Ökosystemen

## Die 3 Säulen des Smart Systems Hubs

---

Integration von Hardware, Software und Connectivity

→ IoT Plattformen

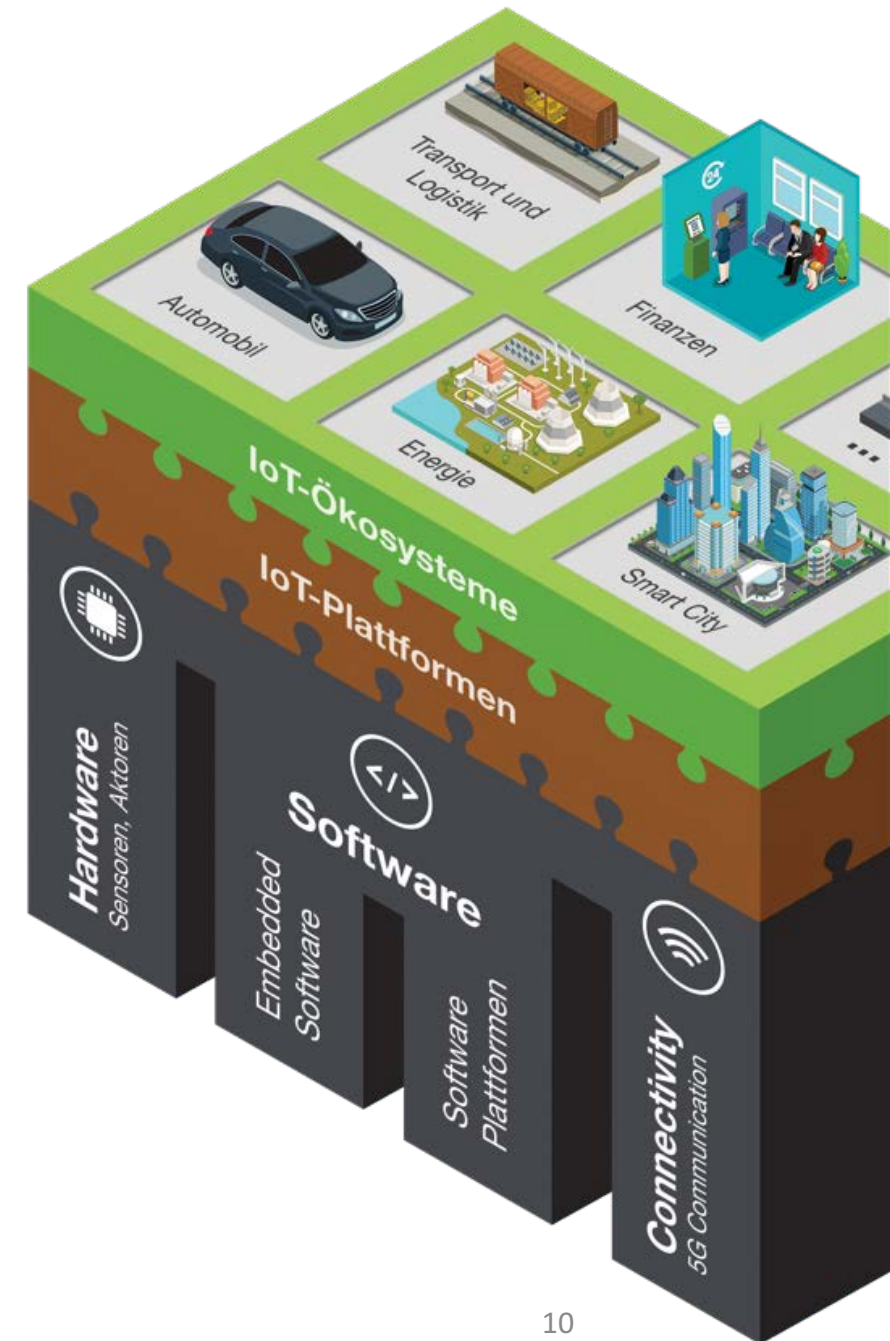
→ IoT Ökosysteme

**Hardware:** Sensoren- und Aktorensysteme (z.B. MEMS)

**Software:** Embedded und High-Level Software als Key-Enabler

**Connectivity:** 5G-Network für IoT-Anwendungen

(Latenz, Datenrate, Resilience ...)





# Der Campus

## Der Smart Systems Hub vernetzt

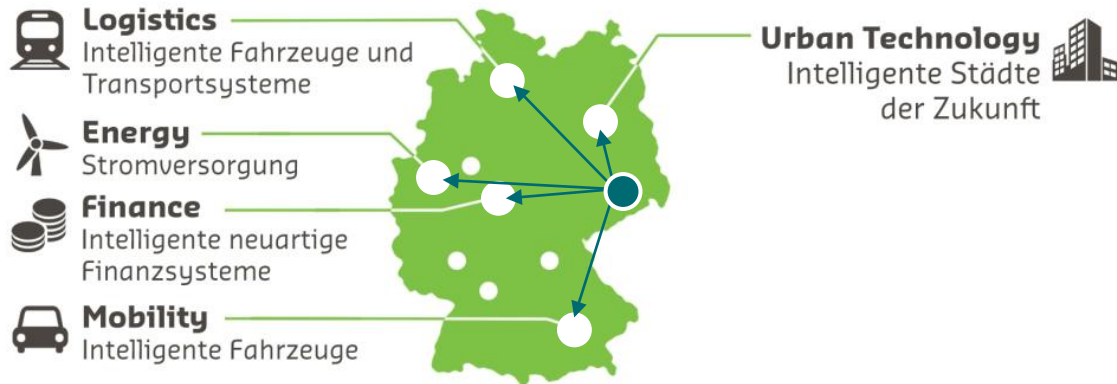
Enges Zusammenarbeiten von  
Forschung, Startups, Inkubatoren, KMUs,  
Flagschiffunternehmen und der öffentlichen Hand

### CIC - Co-Innovation Center (In Planung)

- Neues Gebäude als Zentrum des Smart Systems Hub

### LZB - Lehmannzentrum (In Planung)

- Forschungszentrum des Smart Systems Hub



# Anknüpfungspunkte für das Co-Innovation Center

## Lehmann-Zentrum Phase I: Rechenzentrum + HRSK-II

### Deutscher Rechenzentrumspreis 2014

Sieger in der Kategorie Energie- und Ressourceneffiziente Rechenzentren 2014  
Plenumskonzept des LZR für Energieeffizienz und Sicherheit

- Innovatives Kühlkonzept
- Hohe Energieeffizienz
- Hohe Ausfallsicherheit

### HPC-Installation von Bull

- Baukosten: ca. 45 Mio. Euro
- ~ 44.000 Rechenkerne INTEL
- ~ 1,5 PetaFLOP/s
- ~ 132 TB RAM





# Erweiterung HRSK-II

## Unikat: Infrastruktur für Big Data Analytics (10 Mio. Euro)

### Parallele Speicherhierarchien

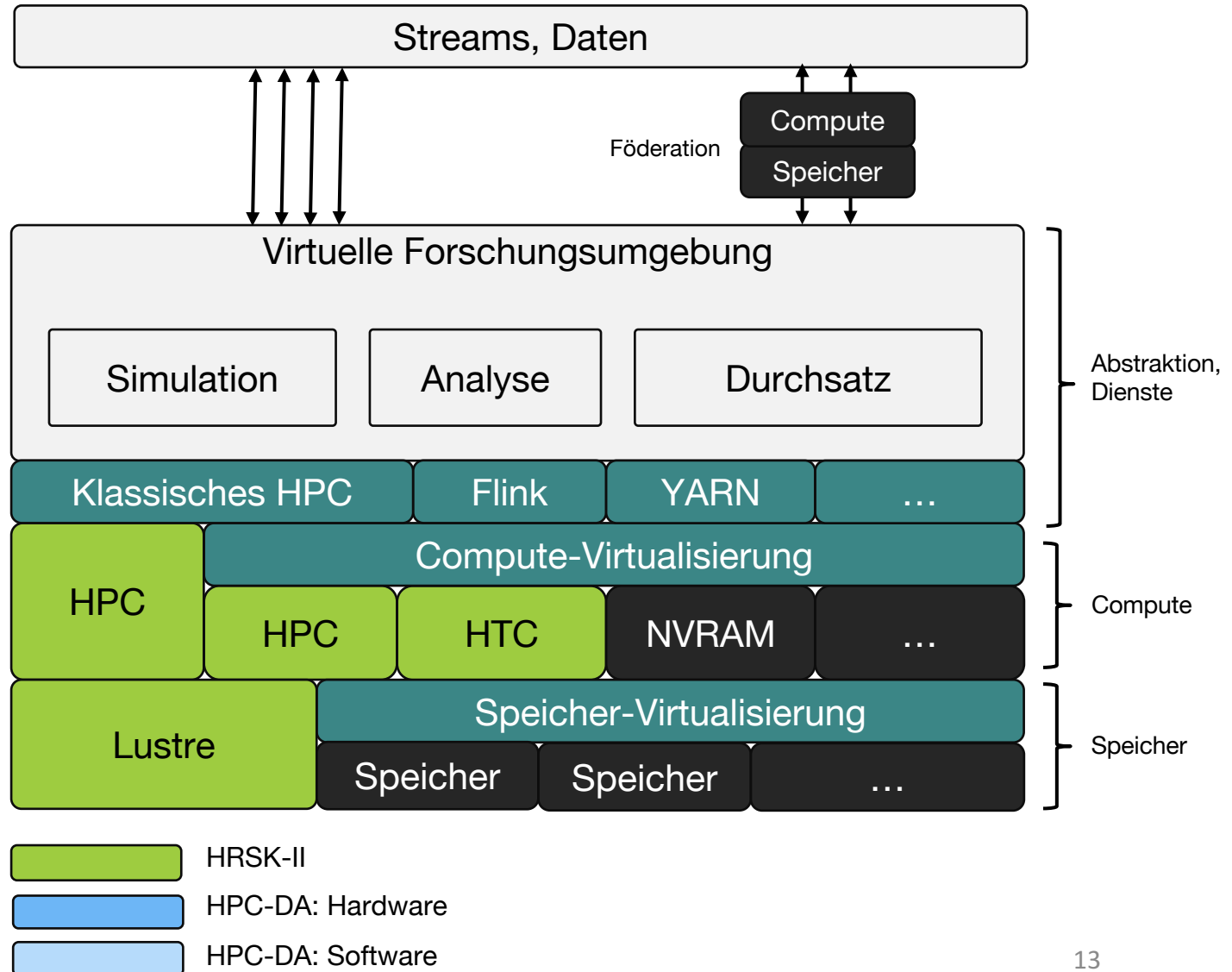
- Tiefere Speicherhierarchien (insbes. I/O-Bereich)
- Effizienter Einsatz von NVRAM

### Data-Analytics-Umgebungen

- Nutzung von Data Analytics- und Big Data Frameworks
- Einbindung in Workflows
- Weiterentwicklung zu Modulen und Diensten

### Virtuelle Forschungsumgebungen

- Entwicklung und Bereitstellung von Basis-Modulen
- Nutzer gestalten daraus ihre Umgebungen
- Umgebungen sollen temporär einfrier-/sicherbar sein



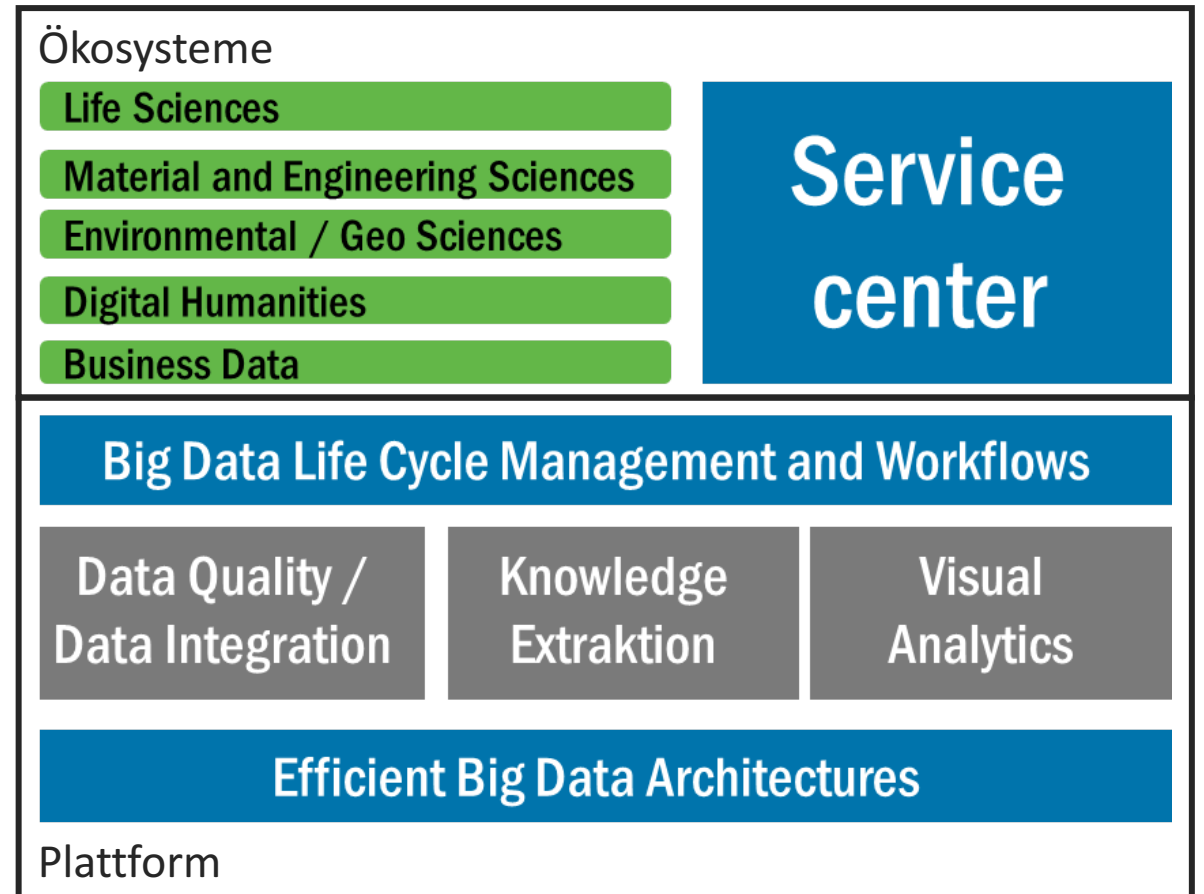
# ScaDS Dresden/Leipzig

## Big Data Kompetenzzentrum (BMBF)

---

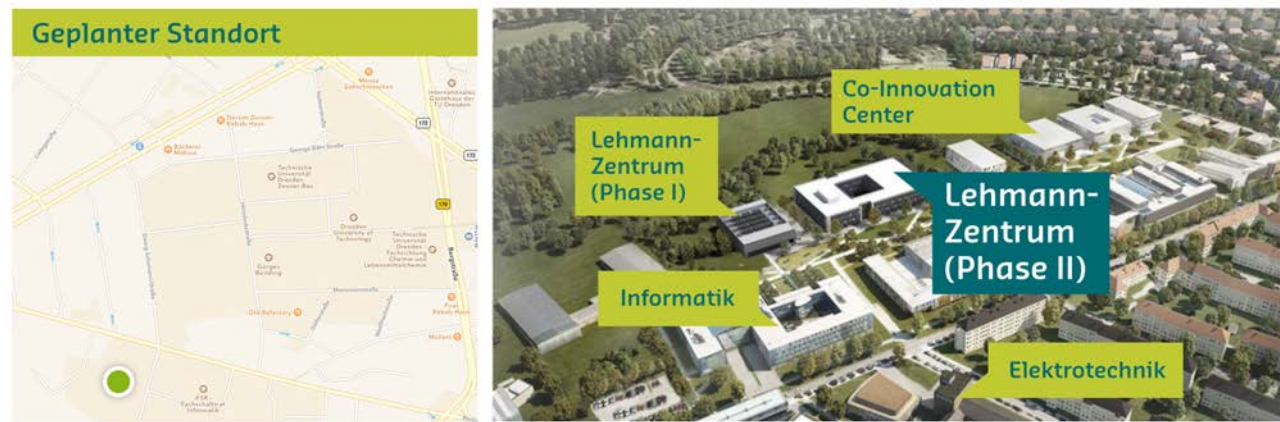
### Entwicklung und Bereitstellung skalierbarer Datendienste für wissenschaftliche Anwendungen über den gesamten Daten-Lebenszyklus (Anwendungsgetriebene Methodenforschung)

- Deep-Learning-Methoden zur semantischen Annotation in Bildern (z.B. Bild- oder Farbsegmentierung)
- Canonical Text Services zur Referenzierung von Textfragmenten und Beziehungsanalyse von Entitäten
- Multilinguale Erkennung von OCR-Daten als Grundlage für semantische Analysen
- Multiskalenvisualisierung zur Korrelation verschiedener Auflösungen in Simulationen und groß-skalige Datenvisualisierung
- Evaluation und Optimierung von Analytics-Frameworks





# Lehmann-Zentrum Phase II (LZB) Organisationen/Themen-Cluster



<b>TEIL 01</b>	<p><b>ZIH</b> Zentrum für Informationsdienste und Hochleistungsrechnen</p> <ul style="list-style-type: none"> <li>• Büros, Seminarräume</li> <li>• Testlabore, Werkstätten, 3D-Labor</li> <li>• Videokonferenzräume, PC Pool</li> </ul>	<b>TEIL 05</b>
<b>TEIL 02</b>	<p><b>LICOSS</b> Living Computer Science Labs Saxony</p> <ul style="list-style-type: none"> <li>• Büros, Seminarräume, Meetingräume</li> <li>• Labore, Werkstätten, Living Labs</li> <li>• Experimentierfläche, IoT Makerspace</li> </ul> <p><b>Teilweise Auslagerung in CIC möglich</b></p>	<b>TEIL 06</b>
<b>TEIL 03</b>	<p><b>LIC</b> Lab für Industrie 4.0 und Cyber-Physikalische Systeme</p> <ul style="list-style-type: none"> <li>• Büros, Projektarbeitsraum, Seminarräume</li> <li>• Labore</li> </ul> <p><b>Teilweise Auslagerung in CIC möglich</b></p>	<b>TEIL 07</b>
<b>TEIL 04</b>	<p><b>MZ</b> Medienzentrum</p> <ul style="list-style-type: none"> <li>• Büros, Seminarräume, Besprechungsraum</li> <li>• Medienlabore, PC Pool</li> <li>• Filmstudios, Audiostudios</li> <li>• Werkstatt</li> </ul>	<b>TEIL 08</b>
	<p><b>ScaDS</b> Competence Center for Scalable Data Services and Solutions</p> <ul style="list-style-type: none"> <li>• Büros, Hörsaal</li> </ul>	<b>TEIL 09</b>
	<p><b>CAMS</b> Center for Advanced Modelling and Simulation</p> <ul style="list-style-type: none"> <li>• Büros, Seminarraum, Meetingraum</li> </ul>	
	<p><b>Verwaltungs-IT</b> SG 6.4. "Datenverarbeitung"</p> <ul style="list-style-type: none"> <li>• Büros</li> <li>• Meetingraum</li> </ul>	
	<p><b>5G Lab Germany</b></p> <ul style="list-style-type: none"> <li>• Büros, Experimentierfläche, Labore</li> </ul> <p><b>Teilweise Auslagerung in CIC oder Barkhausen Institut möglich</b></p>	
	<p><b>Transferzentrum</b> Ausgründungen</p> <ul style="list-style-type: none"> <li>• Büros, Seminarraum, Testlabor</li> </ul> <p><b>Teilweise Auslagerung in CIC möglich</b></p>	

- High Performance und Cloud Computing (ZIH)
- Cyber-Physical Systems, Industrie 4.0 (LIC)
- Internet der Dinge, Sicherheit, Safety, Autonome Systeme, Robotik, Visual Computing (LICOSS)
- Modellierung und Simulation (CAMS)
- Datenintensives Rechnen (ScaDS)
- 5G-Lab Germany
- Medienzentrum
- Technologietransfer

A photograph of a modern building's interior. The space is characterized by a large, curved glass skylight at the top, which allows natural light to illuminate the area. A prominent feature is a large, abstract sculpture in a vibrant green color, resembling a stylized plant or a cluster of rounded forms. To the left, a metal spiral staircase winds upwards. The architecture is clean and industrial, with dark metal frames and light-colored walls. The overall atmosphere is bright and open.

**HERZLICHEN DANK  
FÜR IHRE AUFMERKSAMKEIT!**

**Mit Dank an:**

**Frank Fitzek**

**Uwe Aßmann**

**Christian Piechnick**

**Rene Jäkel**

**Gerhard Fettweis**

**und viele andere Mitstreiter**