



A COMPETENCE CENTER FOR BIG DATA RESEARCH

Dr. Ralph Müller-Pfefferkorn
Technische Universität Dresden
Center for Information Services and
High Performance Computing

www.scads.de

WHY IS BIG DATA A PHENOMENON?

- Cheap and/or easy to use technologies for data taking and computation
 - Web 2.0
 - Internet of Things
 - Industry 4.0
 - Computing (HPC, Grids, Clouds)
- Experiments and simulations producing Big Data
- Access to external and distributed data
- ...

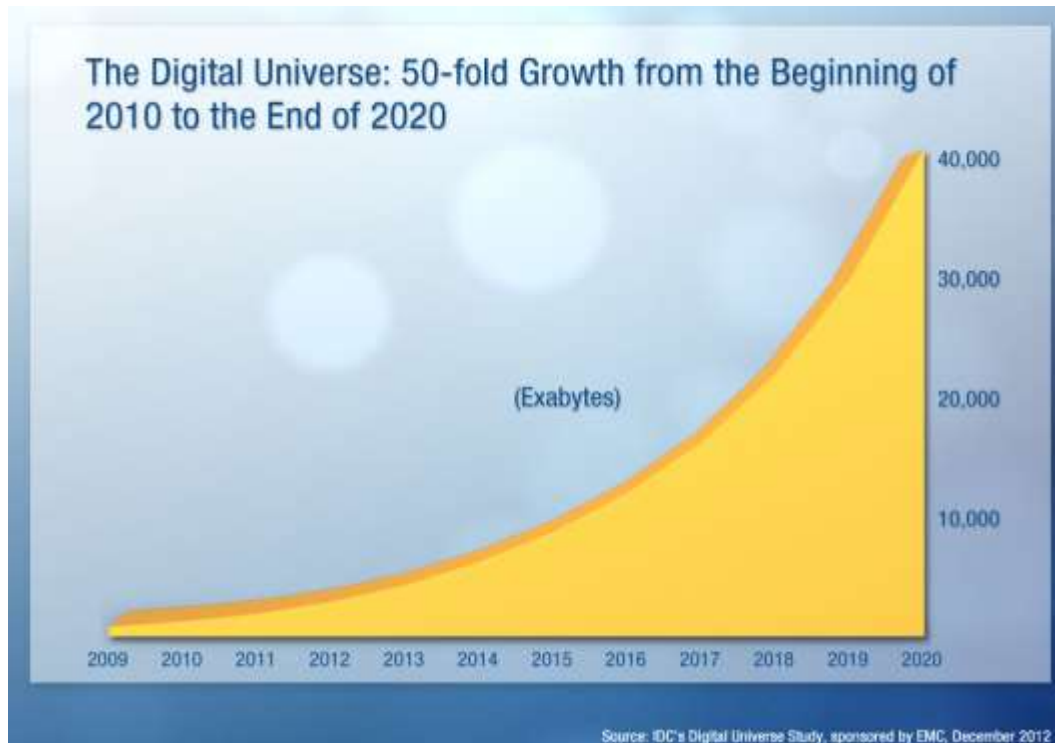


Transformation from
scarce to abundant
data!



Even smaller scientific communities have to use large scale and centralized systems

Gigabytes, Terabytes (10^{12}), Petabytes (10^{15}), Exabytes (10^{18}), Zetabytes (10^{21}) ...



“Big” means
not a fixed
scale!

Need to
adapt and
evolve
permanently.

Current methods for data management and analysis in many scientific domains no longer practical

- Include more and more data from different (distributed) sources in the analysis chain
- Increased conceptual and technical fragmentation and diversification of systems
- Comply various security needs and protection levels of data
- Complex requirements concerning usability (for non IT-professionals) and collaboration of scientists
- ...



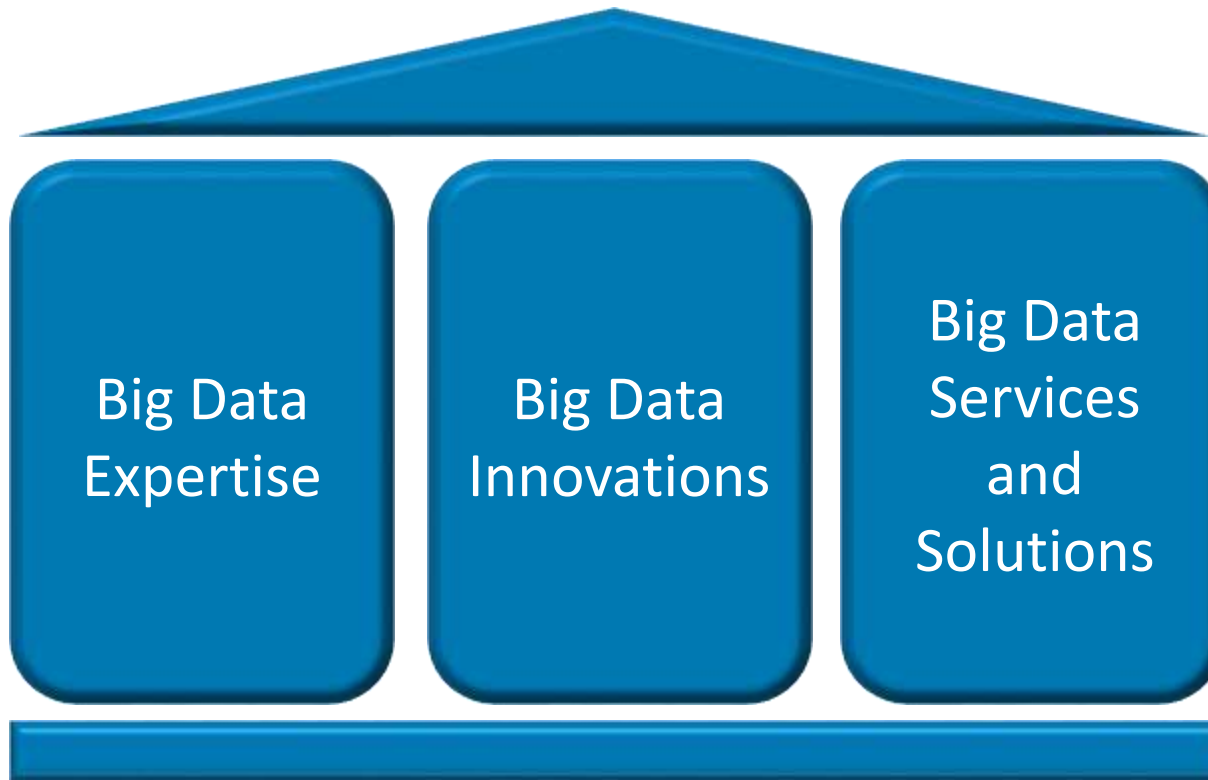
Competence Center for
**Scalable
Data
Services and Solutions
Dresden/Leipzig**





- Project period: 4 years (10/2014 – 09/2018)
- After evaluation possibility of a 3 year extension
- Many research groups involved (>45 PI)
- Computer Scientists and Domain Experts







Leipzig



Dresden

- 16 Fraunhofer institutes
- 6 Max Planck institutes
- 2 Helmholtz centers
- 1 Helmholtz institute
- 6 Leibnitz institutes
- 8 State reasearch institutions
- 14 Institutions of higher education

Technische Universität Dresden



Universität Leipzig



Max Planck Institute of Molecular Cell Biology
and Genetics



Leibniz Institute of Ecological Urban and
Regional Development



Leibniz-Institut
für ökologische
Raumentwicklung

- Avantgarde-Labs GmbH (AL)
- Data Virtuality GmbH (DV)
- E-Commerce Genossenschaft e. G. (ECEG)
- European Centre for Emerging Materials and Processes Dresden (ECEMP)
- Fraunhofer-Institut für Verkehrs- und Infrastruktursysteme (IVI)
- Fraunhofer-Institut für Werkstoff- und Strahltechnik (IWS)
- GISA GmbH
- Helmholtz-Zentrum Dresden - Rossendorf (HZDR)
- Hochschule für Telekommunikation Leipzig (HfTL)
- Institut für Angewandte Informatik e. V. (InfAI)
- Landesamt für Umwelt, Landwirtschaft und Geologie (LfULG)
- Netzwerk Logistik Leipzig-Halle e. V. (NLLH)
- Sächsische Landesbibliothek – Staats- und Universitätsbibliothek Dresden (SLUB)
- Scionics Computer Innovation GmbH (SCI)
- Technische Universität Chemnitz (TUC)
- Universitätsklinikum Carl Gustav Carus (UK)

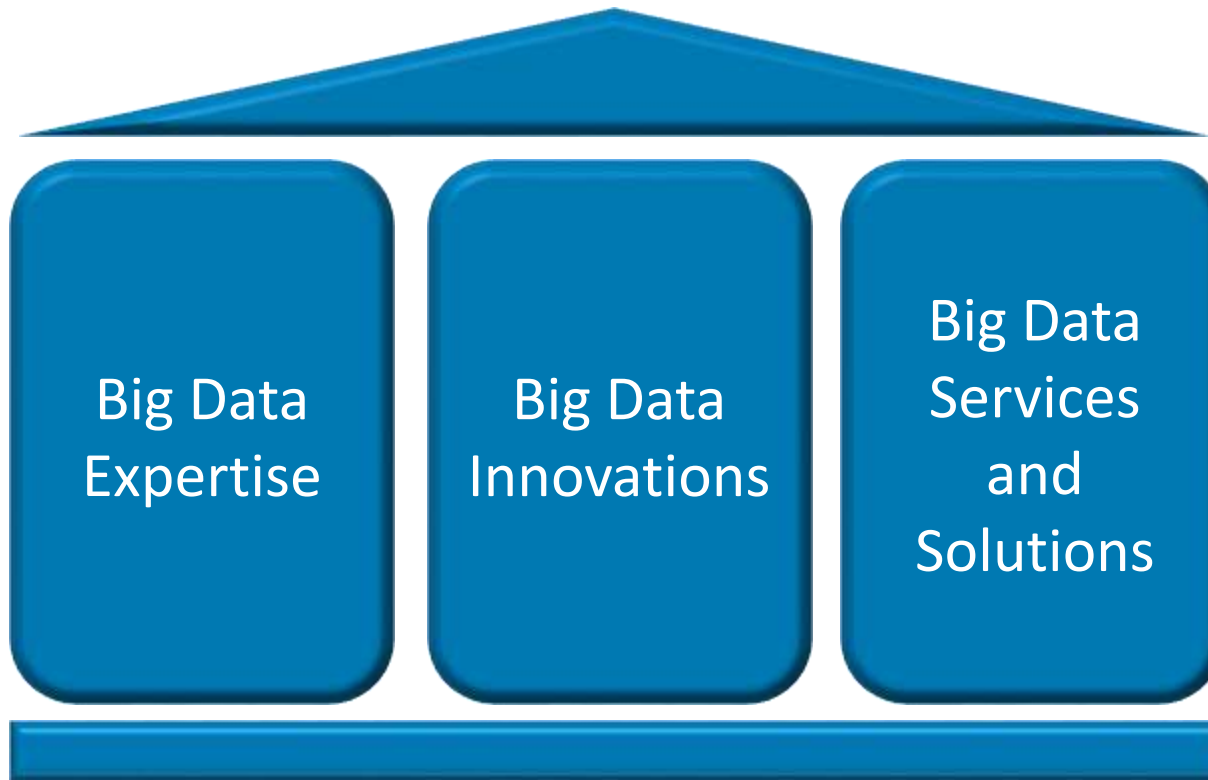
Partners from academia and industry

- Close interdisciplinary cooperation of computer science and user research
- Buildup of an institutionalized network structure
- Knowledge exchange and transfer
 - Training and counselling
 - Teaching
 - Promotion of young scientists
 - ...

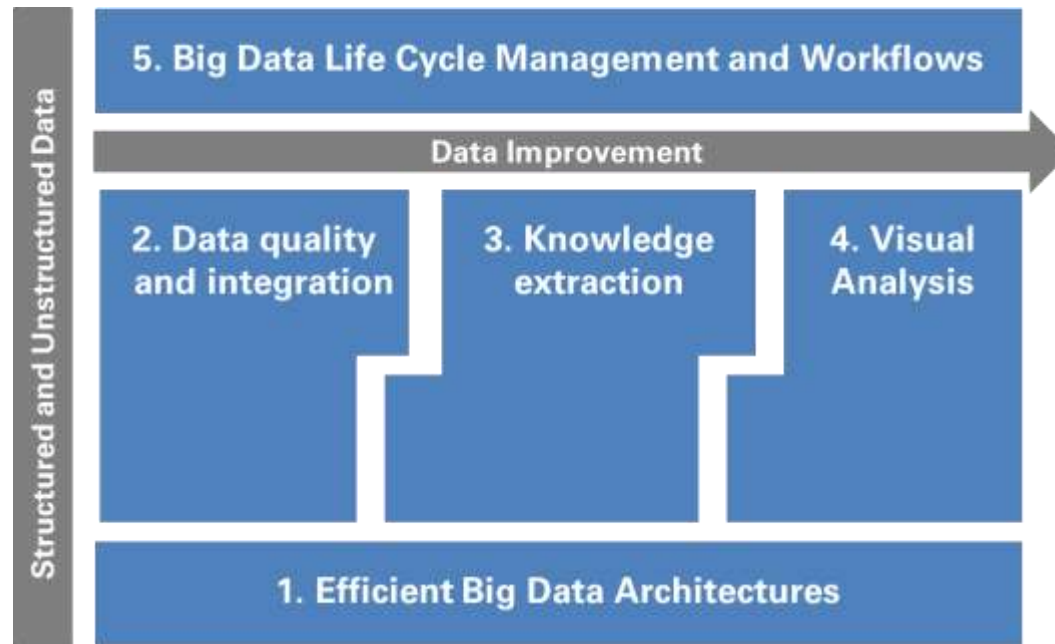


- With other centers of excellence and projects
 - Berlin Big Data Center, BMBF Big Data Projects
 - Competence centers for IT security
 - LSDMA
- Service Center will provide “external” solutions and services
- Joint research
- Common events
 - Topical workshops
 - Conference
 - Summer school ...





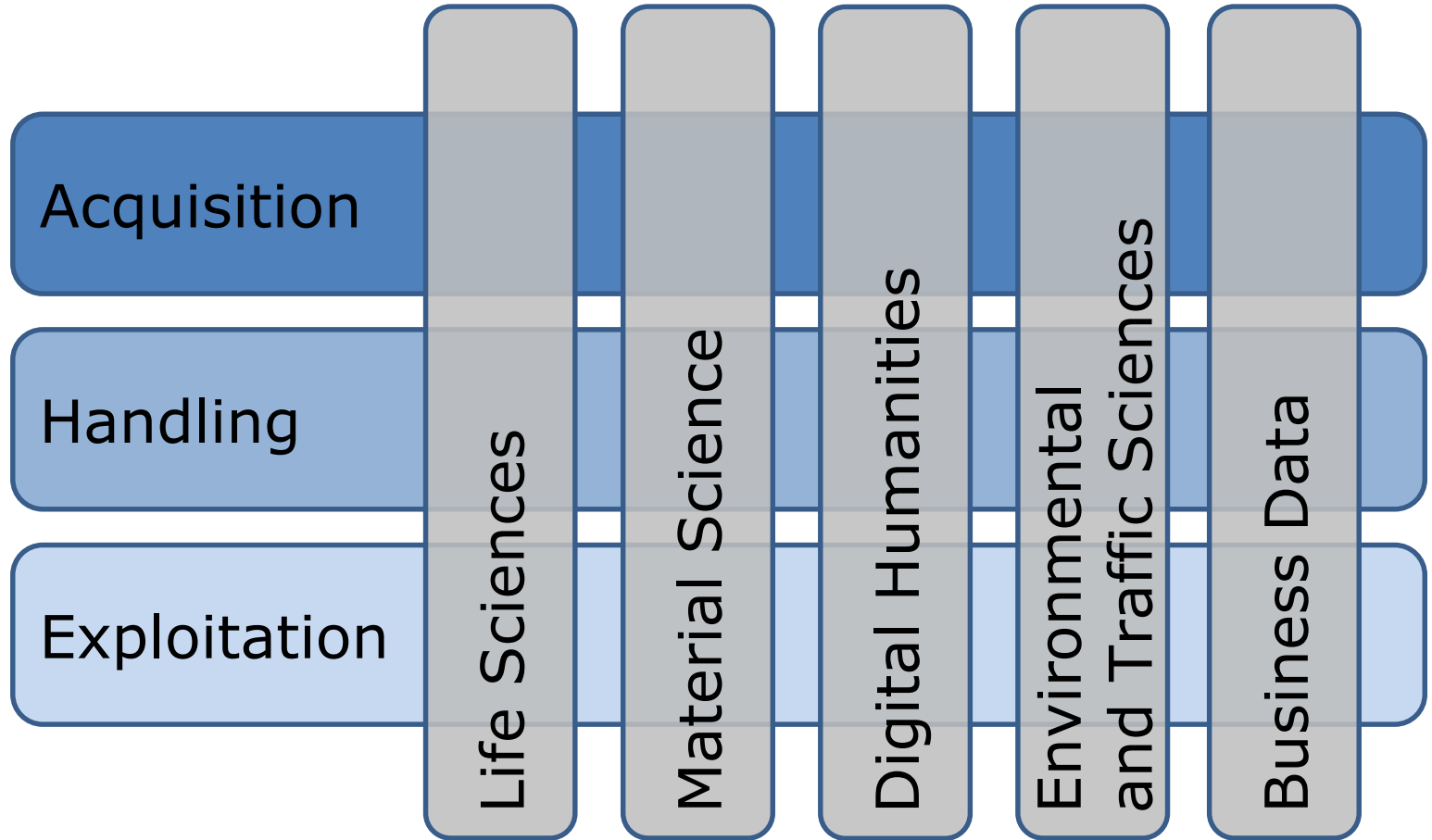
- Excellent and innovative Big Data research
 - Computer Science
 - Application domain specific
- Clear focus on user applications
- Tight coupling of heterogeneous user groups
 - Better abstraction of the basic problem of a challenge
 - Integration of user specifics by direct user contact
 - Cross-disciplinary approaches



- Cooperation on various topics of computer science
- Methodological focus: data quality and integration, knowledge extraction, visual analysis
- Cross-cutting topics: Big Data architectures and data life cycle management

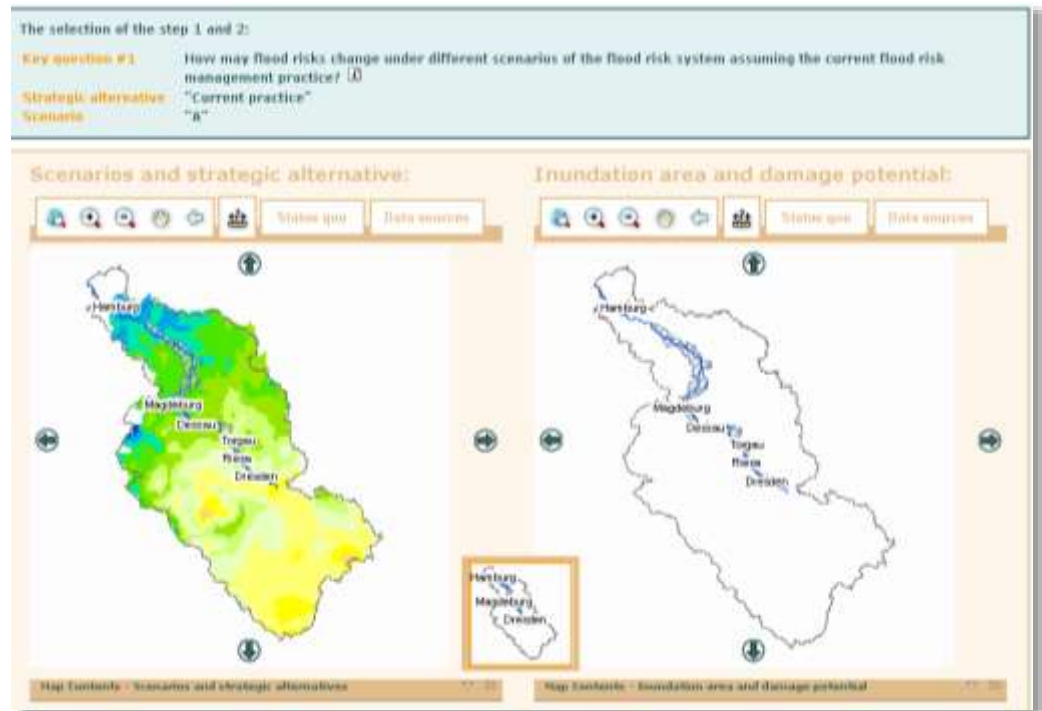
- Efficient Big Data architectures
 - Efficient data compression and intelligent I/O
 - Schedulable data transfer and efficient and (meta) data access
 - Performance analysis of data operations
 - Data security
- Data quality and integration
 - Learning based configuration of integration workflows
 - Parallel execution of comprehensive data integration workflows
 - Real-time data integration and dynamic information enrichment

- Knowledge extraction
 - Efficient algorithms for structural data
 - Machine learning in structural models
 - Text-mining methods for similarity analysis
 - Exploration of metabolic networks
- Visual analysis
 - Alternative reduction techniques and real-time visualization
 - Guided navigation and interactive data validation
- Data life cycle management and workflows
 - Flexible workflows and lifecycle control
 - Data provenance and creation of a relevance model for data



Impact modelling and decision support

- Projections of the multi-model ensemble calculated with numerous impact models
- Parallel and coupled model runs and resulting geo data (> 10 TB)
- Transfer to environment development actors requires accessibility via a web-based spatial decision support system



Local Flood Modelling

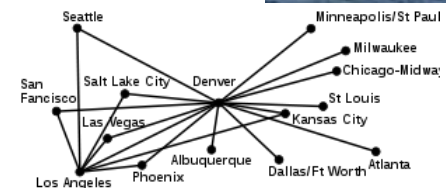
- Transformation of a paradigm
 - Digitization enables access to a vast number of documents (books, manuscripts, papyri, ...)
 - In order to record and analyze content make texts accessible by machines
- Shift from documents to words/phrases/sentences
„Every preserved word (or fragment) in every edition, manuscript, inscription, and papyrus ... is now an object of interest that must possess a unique identifier.“
(Crane et. al. 2012)
- Canonical Text Services (CTS) standard
Example: “urn:cts:demo:shakespeare.sonnets:35.1-35.4”
refers to: Line 1 to 4 of Shakespeares Sonnet 35

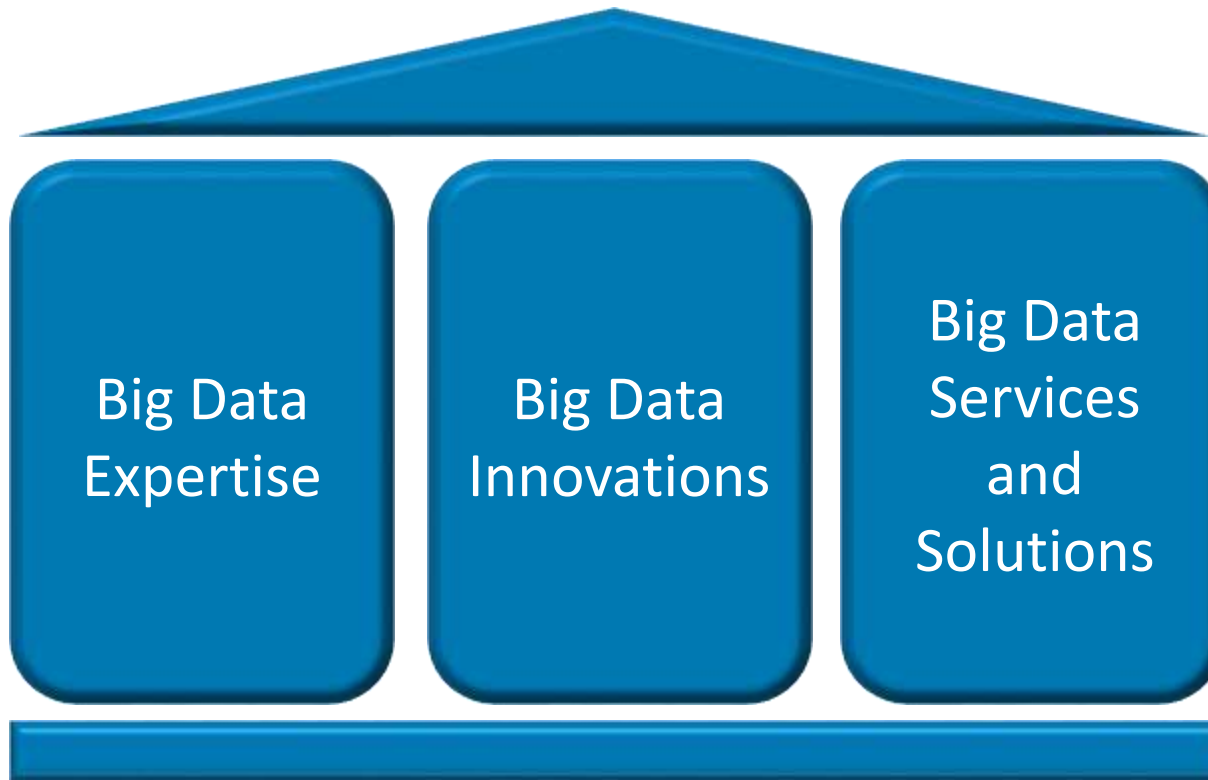
American Standard Version	In the beginning God created the heavens and the earth.
Bible in Basic English	At the first God made the heaven and the earth.
Darby Bible	In the beginning God created the heavens and the earth.
King James Version	In the beginning God created the heaven and the earth.
World English Bible	In the beginning God created the heavens and the earth.
Young's Literal Translation	In the beginning of God`s preparing the heavens and the earth --
New Life Version	In the beginning God made from nothing the heavens and the earth.

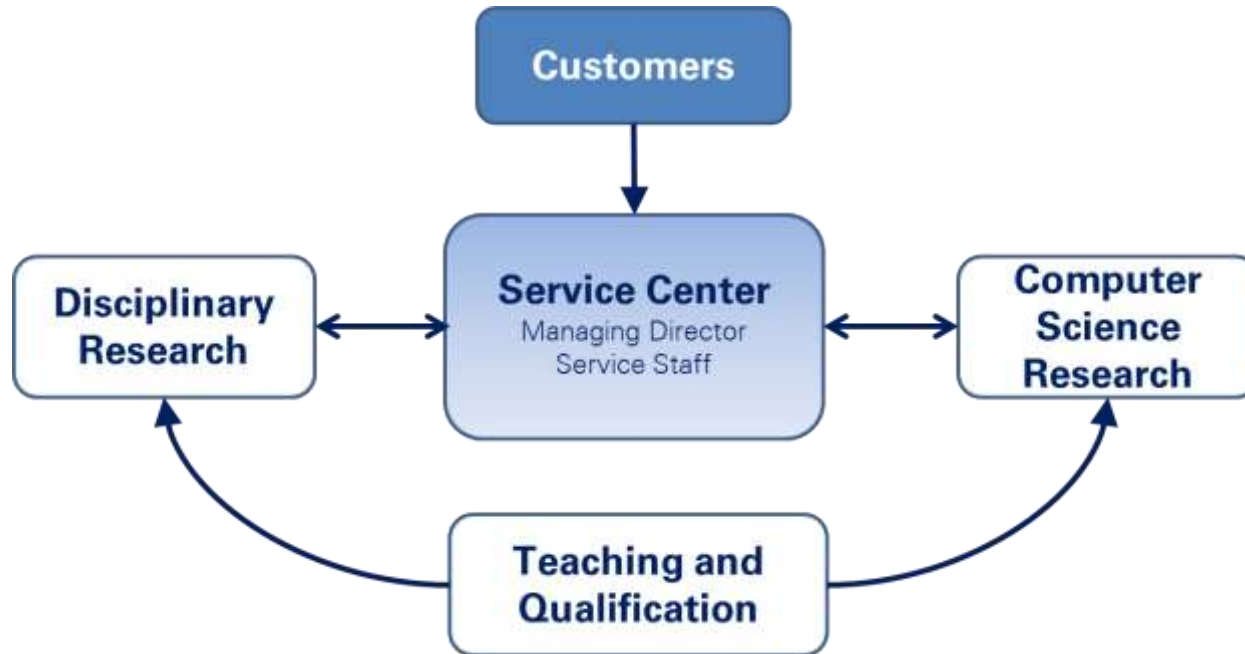


<http://www.informatik.uni-leipzig.de:8080/Bibel/>

- **Synchromodal Transport**
 - Integrated and flexible management
 - Using all available means of transportation
- **Wide variety of data sources**
 - Real-time and historic data
 - Internal and external data
 - Traffic and goods data
- **Challenges**
 - Data integration and quality
 - Real-time analysis







- Service Center as a "Single Point of Contact"
- Formal and structural organization of (distributed) research and development
- Synergetic consolidation and organization of the knowledge transfer

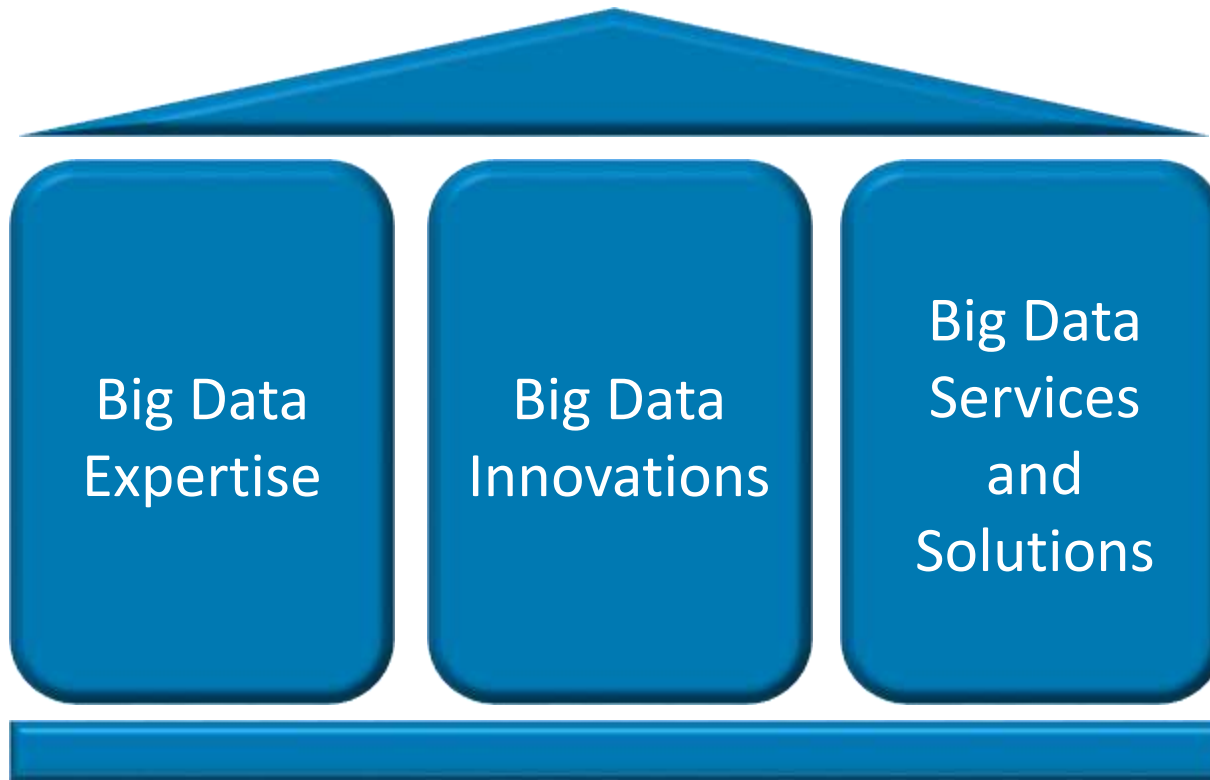
- Provide
 - Own and external solutions
 - Services

- Modular and service-oriented approach
 - Aligned with the needs of the users
 - Cross-disciplinary service portfolio
 - Problem oriented concepts and solutions
 - Operation and support

- Big Data is a reality in science.
- Not an unique solution for all, but different aspects for heterogeneous challenges.

Support scientific users in their daily work!

**Provide infrastructures and
methods for their specific needs to handle
Big Data.**



SCIENTIFIC COORDINATION

Prof. Dr. Wolfgang E. Nagel

Technical University Dresden

Center for Information Services and High
Performance Computing
01062 Dresden

Phone: +49 351 463-35450

E-Mail: wolfgang.nagel@tu-dresden.de

MANAGING DIRECTOR

Dr. René Jäkel

Technische Universität Dresden

Zentrum für Informationsdienste und
Hochleistungsrechnen
01062 Dresden

Telefon: +49 351 463-42331

E-Mail: rene.jaekel@tu-dresden.de

Prof. Dr. Erhard Rahm

Universität Leipzig

Fakultät für Mathematik und Informatik
Augustusplatz 10
04109 Leipzig

Phone: +49 341 97-32221

E-Mail: rahm@informatik.uni-leipzig.de