DFG – Commission on IT Infrastructure (KfR)

Study Tour 2013 – Report

Deutsche Forschungsgemeinschaft
Kennedyallee 40 · 53175 Bonn · Postanschrift: 53170 Bonn
Telefon: + 49 228 885-1 · Telefax: + 49 228 885-2777 · postmaster@dfg.de · www.dfg.de

General Remarks

The DFG Commission on IT Infrastructure (KfR) makes recommendations on general issues of IT infrastructure within the German academic and research environment. To do so, it has proven a good practice to conduct bi-annual study tours to countries abroad, in order to gain insights into latest developments and best academic practices in IT services and IT infrastructures at leading research laboratories and universities. In addition, the commission takes into account IT trends as they are envisioned by the IT industry as well as relevant funding schemes of the respective national funding agencies. The discussion with both these stakeholders, industry and funding organizations, is essential for reflecting on the framework of constraints and needs of the German academic landscape.

Issues at the core of the KfR's 2013 study tour were cloud services, approaches to IT governance, IT infrastructures, the emerging field of Computational Science and Engineering, High-Performance Computing (HPC), life science research, medical services, and Big Data. To enable the Commission to receive a broad level of expertise in these areas, the 2013 Study Tour focused on research and higher education institutions, companies, and funding agencies in the U.S., since many of the actors driving the development of the topics mentioned above are located there. Furthermore, the Commission received additional expertise through a related visit to the CERN facility and the EPFL in Switzerland.

Outcomes

The following remarks reflect the most important findings of the KfR Study Tour 2013 in a condensed form. These findings will either be included as input for the next KfR recommendations (to be published autumn 2015) or – in more urgent cases, such as caveats related to the use of cloud services – be disseminated by means of an addendum to the current recommendations 2011–2015.

Cloud Services. Cloud services have seen a rapid development in the last years. Commercial providers offer a variety of solutions and services, and at the same time, private clouds are emerging within or across public institutions. A further gain in importance can be expected for virtualization strategies, such as operating a larger number of (logical) servers with special functions (email, web) or for various groups on one (physical) machine.

Commercial clouds can be an interesting alternative to the procurement of IT hardware or even the use of central compute center facilities. However, there are a couple of caveats and risks which have to be considered in each specific case. Among these are security, safety, and privacy, especially for personal and medical data; legal and social/acceptance issues;

questions of ownership (concerning data, licenses, and intellectual property stored in a cloud); long-term dependencies (feasibility of a later provider change or insourcing with respect to time and cost); cost at total-cost-of-ownership level. Especially concerning the latter, it was interesting to observe that economy-of-scale effects let internal/private cloud solutions very often appear more cost-effective than commercial ones. Actually, best practices of IT outsourcing in an academic or research context seem to be limited to rather small or specialized institutions (the Broad Institute in Boston, e.g.) – and also here IT outsourcing entails high costs.

On the other hand, internal/private cloud solutions are discussed in many places. While the University of California system has launched IT structures (such as a system-wide Chief Information Officer, CIO) to establish system-wide services across the (very independent) single universities, a different approach can be seen in North Carolina. Here, the North Carolina State University in Raleigh developed a rather comprehensive and elaborate university IT cloud (where, just as an example, a number of compute nodes can be easily booked in advance for lab courses through a central system), in close partnership with a big IT company. The model is currently rolled out to the NCSU system, with other universities and public institutions expected to join. As another example, the San Diego Supercomputing Center offers secured storage services for a health organization separated from its main cluster infrastructure.

Recommendations: Having in mind these developments, key players in the German academic landscape should be encouraged to survey private cloud strategies at various levels – comprising different institutions at one location, at state-level, or nation-wide. Such solutions allow for both benefitting from economy-of-scale effects and significantly reducing the risks (security, privacy) and possible high long-term costs (migration, intellectual property) of commercial clouds. Some locations already have established concepts (Göttingen, München, or Karlsruhe, e.g.) and in some states, currently strategies are developed and implemented (Baden-Württemberg for HPC and Data and North Rhine-Westphalia for Data, e.g.). On a nation-wide scale, the current preparation of online storage cloud services within the German National Research and Education Network (DFN) has to be mentioned here. Funders should support these directions by adequate schemes and create appropriate environments for these kinds of strategies.

Concerning funding, we see the necessity for changes of paradigm in the general thinking and in the legal framework in Germany. There is an obvious trend from hardware-dominated systems via hardware- and software-dominated systems to an increasing presence of services. Hence, there should be ways to get services funded as well as investments. Moreover, given the significant operational costs associated with large IT infrastructures, new *ways of considering investment vs. operational costs should be discussed. In any case, the total-cost-of-ownership concept defining the overall cost including procurement and operation of systems should be accepted as the adequate measure for assessing the costs to run an IT infrastructure.*

IT Governance. As a general observation, the broad variety of IT governance models and, in particular, of CIO positions seems to be similar in scope in the U.S. and in Germany. However, two things should be noted: First, in the U.S., IT is seen as a strategic and vital topic for an academic institution, more than it is in Germany – a key factor for the success of U.S. universities – and this is reflected by organizational structures, allocated human resources, and budgets. In Germany, even for the very cost- and performance-intensive IT at university medical centers, organizational structures including a CIO and state-wide harmonized concepts are rather rare. Second, universities that significantly modernize their IT organization and governance can harvest significant benefits in terms of improved processes, better support for research and education, and, thus, a higher level of attractiveness for faculty and students.
Universities in the U.S. look for and find positive economic effects as one outcome of a clearly designed and vividly implemented IT strategy, especially on a longer time scale. However, short-term savings can and must never be the dominant driving factor – one important core goal is the improved and sustainable quality of services.
Furthermore, there are similar crucial problems in the U.S. as there are in Germany, such as an increasing dependency on software monopolists and a foreseeable hike in software licensing cost. There also seem to be no answers to these questions in the U.S.

Recommendations: All German universities including medical departments should be encouraged to implement appropriate IT concepts. Currently, an increasing gap can be observed between universities addressing these challenges and a still larger number of institutions showing a lot of deficits on that behalf. This has already been described in the KfR recommendations 2011–2015, but the necessity to go into this direction is even more evident nowadays. An IT strategy should be established at the management level of each university to foster the respective overall university strategy. IT has to be recognized as an important driver and enabler of innovations. The university IT concepts should then be integrated in state-wide (as seen in California and as already launched in Baden-Württemberg, e.g.) or even national IT strategies (as they have been developed in the eighties for the network area (cf. the German Research and Educational Network, DFN) and later for HPC).

Computational Science and Engineering. CSE has become a key technology for science and industry. Computer-supported or computational work is ubiquitous in most scientists' life, especially in science and engineering – just think about how ubiquitous computations, large data sets, and data exploration have become in processes in both science and industry. In the U.S., many strategic reports have emphasized the increasing relevance of CSE as well as its inadequate anchoring in the disciplinary landscape. Yet the situation is still diverse. On the one hand, there are extremely successful examples, such as the Scientific Computing and Imaging Institute (SCI) at University of Utah at Salt Lake City or the Institute for Computational Engineering and Sciences (ICES) at University of Texas at Austin, which are widely independent trans-disciplinary institutions (thus acquiring the status of a new disciplinary structure). But these are rather an exception to the rule, where CSE-related researchers are spread across the traditional discipline and department structure (in applied mathematics, computer science, engineering, natural sciences, life sciences, …). The picture concerning funding is similar: While the Department of Energy (DoE) basically has its own dedicated CSE sub-division (Advanced Scientific Computing Research, ASCR), the National Science Foundation (NSF) uses its Office for Cyber-Infrastructure (now the Division for Advanced Cyberinfrastructure within the Computational Science and Engineering Directorate, CISE) as a cross-sectional division, where different directorates/disciplines can use matching funds when funding CSE-related activities.

Recommendations: It is obvious that CSE is an important topic area which is highly interdisciplinary and where is significant need and value to foster research and education. Consequently, and independent of the organizational frame, CSE needs funding for research and education, and there are several large-scale programs, in particular in the context of HPC (cf. the DoE's SciDAC program – Scientific Discovery through Advanced Computing, where cross-disciplinary and distributed topical centers for CSE- and HPC-related research have been established all over the country and the academic landscape) that are worth a deeper consideration for the development of funding schemes in Germany.

High-Performance Computing. While there is a visible trend for organizing computing as a, to some extent, virtualized internal service (within a university, e.g.), commercial compute clouds do not play a significant role in academic environments, apart from, maybe, mere routine computations. As a result, the discussion on cloud computing has become much more centered around storage or general IT services (email, software-as-a-service, etc.) than around classical computing. Real HPC in the cloud is currently hardly an option.

Recommendations: Concerning funding, HPC projects increasingly comprise consortia of various institutions jointly addressing a specific topic – thus integrating compute centers, *HPC methodology specialists (informatics and mathematics), and application domain specialists (cf. the US SciDAC or xSEDE programs), also covering educational aspects. It can also be observed that funding of HPC systems of the highest classes is frequently accompanied by such "usage-oriented" research programs, taking into account that hardware also needs software, and both need "brainware", i.e., HPC expertise and its availability for domain scientists. Funding programs should include appropriate measures to address these trends. DFG's Priority Program 1648 "SPPEXA – Software for Exascale Computing" is an important step in that direction.*

Big Data. There was an interesting consensus among the experts in the discussions that data-driven science and engineering is a recent logical specialization of CSE, but not a completely new paradigm ("4th pillar") as it is frequently promoted. While CSE was originally seen as a synonym for simulation-based science and engineering, the rising data issue shows that "computational" is more multi-faceted.

Nevertheless, it is obvious that the data issue has significant impact. First, data acquisition, compression, storage, archiving, retrieval, mining, analysis, or exploration have all become increasingly standard components of computational work – with significant ramifications related to the need for experts, education, funding, etc. Second, the data topic is a "CSE door opener" for many disciplines, including those at the so-called "long tail" of scientific data with less of a traditional link to the field – again, with all consequences this development entails. And third, size does matter, i.e. the "big" (if speaking of really huge data sets), is related to a variety of challenges such as communication, networks, high-throughput computing, distributed storage, etc. – creating an increasing need for large-scale data facilities and networks. Furthermore, Big Data does not only mean huge volumes. This term is often related to the four characteristics volume, velocity, variety, value. Especially in medicine volume and variety today come together. The establishment of electronic health records on one side and high-throughput genome sequencing technologies on the other side offers the opportunity to integrate phenotype patient data with molecular and genomic data of immense volumes in order to enable translational and personalized research in medicine. Evidently, for those challenges highly interdisciplinary collaboration between clinicians, basic researchers in medicine, medical informatics experts, and computer scientists is indispensable. Finally, the handling of data bases with personal sensitive data (genome-derived information or the Utah population database, e.g.) seems to be considered less critically by the U.S. public than in Germany. Thus, in Germany we face the challenge for similar interdisciplinary and innovative work, but have to further consider our high requirements on data and privacy protection.

Another interesting aspect is balancing generic approaches and domain-specific solutions. Talking to domain scientists, there is the predominant opinion that the needs of the domain science in terms of data formats etc. are so specific that each one needs its own solutions. On the other hand, computer scientists tend to aim at generic and general-purpose solutions that can easily be adjusted to specific needs (which often appears to be the more elegant and efficient way). This raises the question whether we need data technologies and data centers for each field, or to what extent and under what circumstances such an approach might be reasonable. This is currently not completely clear – but some coordination beyond single applications should be mandatory. And, of course, the overall challenge is still to make scientists in all fields really use the available data technologies in a standardized and domain-appropriate way.

Recommendations: *Big Data (which, interestingly, almost always needs "big computing" to extract the science and knowledge out of the data) has certainly emerged as a topic to be addressed in the next KfR recommendations. The BMBF has already launched calls related to Big Data, a respective DFG Priority Program has also just been announced. Universities and IT service centers have to consider the above-mentioned consequences – both from a technical point of view and from specific scientific requirements coming from the respective disciplines. The newly established DFG program on Research Data might be used by research communities to address their needs with respect to an appropriate handling of research data in general and large data sets in particular. Moreover, research data are of importance for state-wide IT concepts.*

Concluding Remarks

This report is meant as a high-level summary of KfR's recent study tour which actually was complemented by a regular meeting held in Switzerland (CERN, EPFL). As initially stated, an addendum to the regular KfR recommendations is envisaged for the end of 2013 / beginning of 2014, especially since there is an obvious need for advice on how to proceed in areas such as cloud computing – for individual scientists as well as for the DFG as a funding agency.

The DFG is highly recommended to address the definition and scope of large instrumentation funding programs. The current situation, i.e. the sharp distinction between investment costs (applicable via Art. 91b GG or Art. 143c GG derived funding schemes) and operational costs (to be covered by the host institutions) will run into severe problems, especially with IT infrastructures. Since the boundary conditions are set by Joint Science Conference (GWK), it should be discussed how to address the developments from DFG's point of view.

KfR Study Tour 2013, March 03-16

Core topics

- Cloud services – demands, offers, best practices, chances & risks
- IT governance – structures, organization, implementation, strategic relevance
- Computational Science & Engineering (CSE) – a discipline?
- High-performance computing (HPC) – organization, funding
- (Big) data – consequences & needs of a data-driven science

Institutions visited[1]

- Seattle:
  - Microsoft
  - University of Washington, Medical Center
- San Diego:
  - University of California at San Diego (UCSD), UC system, IT governance
  - UCSD, San Diego Supercomputing Center (SDSC)
- Salt Lake City:
  - University of Utah, Scientific Computing and Imaging Institute
  - University of Utah, Huntsman Cancer Institute & University hospital
- Washington D.C.:
  - National Science Foundation (NSF), Office for Cyber-Infrastructure (OCI)
  - Dept. of Energy (DoE), Advanced Scientific Computing Research (ASCR) division, Scientific Discovery through Advanced Computing (SciDAC) program
  - National Institutes of Health (NIH), National Library of Medicine, Biomedical Translational Research Information Systems (BTRIS), National Center for Biotechnology Information
- Washington D.C. – Workshop at the DFG North America Office
  - National Academy of Science, Board on Research Data and Information (BRDI), Computer Science and Telecommunication Board (CSTB)
  - Coalition for Networked Information (CNI)
  - Indiana University, data-driven science, IT governance
  - DARPA
- Raleigh:
  - North Carolina State University and IBM – cloud services
- Boston:
  - Broad Institute

[1] In February 2013, the KfR held a regular meeting in Geneva, Switzerland, combining it with site visits at CERN, Geneva (IT governance, Big Data) and at EPFL, Lausanne (IT governance, CSE). Observations and findings from these visits have given additional input to this report.

Participants

KfR members

- Prof. B. **Awiszus** (TU Chemnitz)
- Prof. J. **Bernarding** (U Magdeburg)
- Prof. H.-J. **Bungartz** (TU München, Head)
- Prof. O. **Kao** (TU Berlin)
- Prof. P. **Loos** (U Saarbrücken)
- Prof. C. **Marian** (U Düsseldorf)
- Prof. U. **Prokosch** (FAU Erlangen-Nürnberg)
- Prof. R. **Yahyapour** (GWDG Göttingen)

Additional experts

- Prof. C. **Bischof** (TU Darmstadt)
- Prof. W. E. **Nagel** (TU Dresden)

DFG head office

- Dr. J. **Janssen** (WGI)
- Dr. S. **Winkler-Nees** (LIS)
- Dr. M. **Wilms** (WGI, Organization)

DFG North America office

- Dr. M. **Vögler** (DFG-Washington)