

# Challenges for Future Platforms, Services and Networked Applications

Krzysztof Kurowski

Poznan Supercomputing and Networking Center,  
Noskowskiego 10, 61-704 Poznan, Poland  
krzysztof.kurowski@man.poznan.pl  
<http://www.psnc.pl>

**Abstract.** The keynote will address various ICT aspects related to future platforms, applications and services that may impact and hopefully improve the way various business and enterprise processes will be organized in the future. New development tools, easy-to-program powerful mobile devices, interactive panels, etc. enable users to prototype new IT systems with much better interfaces, to quickly setup collaborative environments and easily deploy online applications. However, many people have not realized how quickly data, computational and networking requirements of these new information systems have increased and what constraints are behind existing underlying e-Infrastructures. Therefore, the main aim of this keynote will be to share with the audience example scenarios and lessons learned. Additionally, various challenges encountered in practice for future information systems and software development methods will be briefly discussed. The keynote will be summarized by an updated report on existing research e-Infrastructures in Poland and worldwide.

**Keywords:** Future Internet, Grid Computing, Cloud Computing, Data Management, Multimedia.

## 1 Introduction

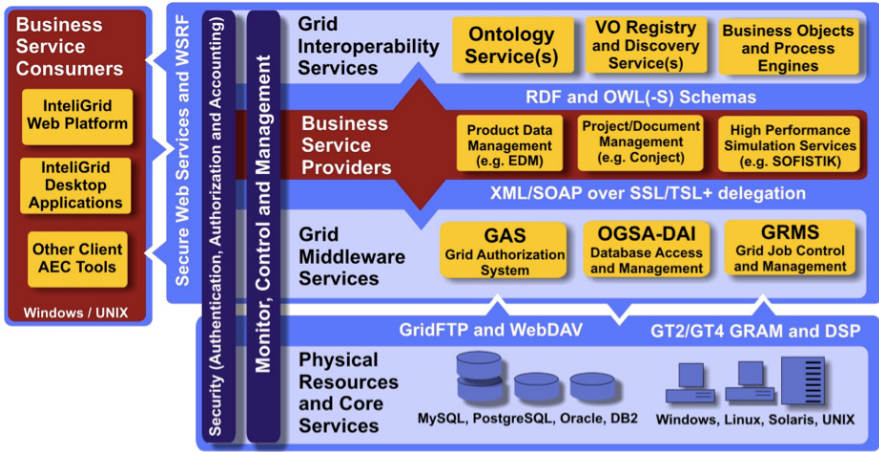
Grid computing has been promoted for more than ten years as the global computing infrastructure of the future. Many scientists and others have considered Grid computing as one of main sources of the impact that scientific and technological changes have made on the economy and society [18]. This claim is based on the observation that the usage of large data volumes has become increasingly important to many disciplines, from natural sciences, engineering to the humanities and social sciences. However, despite significant investments in the grid concept, the number of users is not increasing. Instead, new concepts (or at least new terms) like Clouds seem to be replacing the grid computing approach (or name). Whatever is said about grid computing, it is still a key element of many global e-Infrastructures. Today, the largest scientific computational collaborations have deployed and depend on grid computing infrastructures as their

production computing platforms. For instance, the high energy physics detector ATLAS at the Large Hadron Collider (LHC) is recording data which is distributed within the Worldwide LHC Computing Grid (WLCG) consisting of more than 140 computing center in 35 countries and the 4 LHC experiments [21]. Fusion research is concentrated on studying the properties of plasmas, which require a large computing effort that is provided by different grid environments [6]. Moreover, the grid concept has affected design and development processes of many innovative IT systems over the last decade, and two of them are briefly introduced in this paper together with various challenges for future applications and services delivered on future e-Infrastructures.

The rest of the paper is organized as follows. In Section 2, the IntelliGrid project is presented as a good example how grid technologies can be used to address the lack of integrated computing environment for engineering industries. The IntelliGrid case study shows challenging integration and interoperability needs with a flexible, secure, robust, ambient accessible, interoperable, pay-per-demand access to information, communication and processing provided by the e-Infrastructure. Then, Section 3 shows interesting case study aimed at the development of open-source, semantic and grid-based technologies in support of post genomic clinical trials in cancer research. In Section 4, some key computational challenges in the use of emerging hardware architectures for future applications are presented. Section 5 introduces another challenging problems related to an increasing number of users, digital devices and services delivered via the high-speed Internet and Next Generation Networks. Finally, Section 6 concludes the paper by defining some challenges for future e-Science e-Infrastructures.

## **2 The IntelliGrid Project: e-Infrastructure for the Architecture, Engineering and Construction Industry**

In addition to large data volumes distributed processing, one of the key ideas introduced by Grid computing in the late nineties was the concept of virtual organization (VO). VO is perhaps best understood through the following: "... the sharing that we are concerned with is not primarily file exchange but rather direct access to computers, software, data, and other resources, as is required by a range of collaborative problem solving and resource brokering strategies emerging in industry, science, and engineering. This sharing is, necessarily, highly controlled, with resource providers and consumers defining clearly and carefully just what is shared, who is allowed to share, and the conditions under which sharing occurs. A set of individuals and/or institutions defined by such sharing rules form what we call a virtual organization" [8]. Over the last decade, project-oriented modes of operation were also becoming ubiquitous in many industrial sectors as an attempt to provide both increased flexibility and agility to operations, and one-of-a-kind products and services were becoming a norm rather than an exception. For instance, building construction or facilities management increasingly required the one-time collaboration of different organizations to consolidate



**Fig. 1.** The concept of Virtual Organization with example enterprise and business applications integrated through the distributed InteliGrid platform

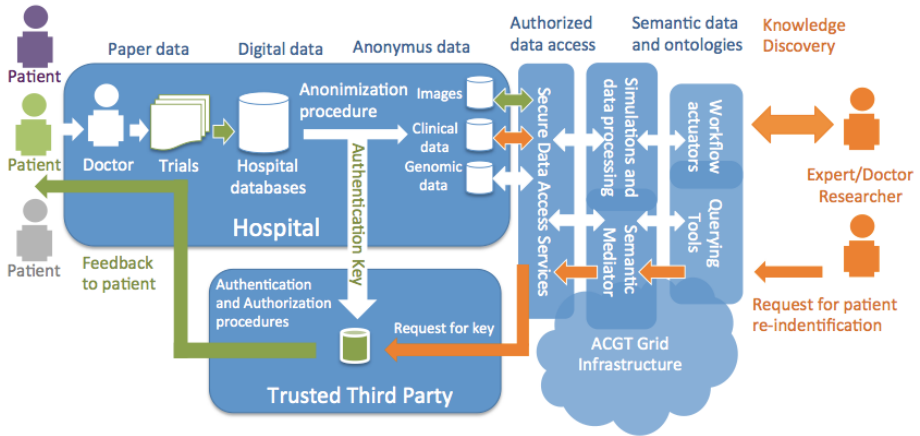
and synergize their dispersed competencies in order to deliver a desired product or service. This naturally had an implication not only on the way information (related to the to-be-delivered product or service) was exchanged and shared, but also on the way in which secure, quick to set-up, transparent (to the end-user) and nonintrusive for day to day work of an individual/organization. Thus, ICT was naturally used for this purpose. Consequently, the VO concept defined as geographically distributed, functionally and culturally diverse, dynamic and agile organizational entities linked through ICT has been a central research theme in various research initiatives [12]. The EU-funded project, InteliGrid (2004-2007) was meant to be a bridge between differing research communities and made available Grid solutions that are usable by VOs as seen by the Architecture, Engineering and Construction (AEC) industry [10]. In a nutshell, InteliGrid combined and extended the state-of-the-art research and technologies in the areas of semantic interoperability, virtual organizations and Grid computing in order to provide diverse engineering industries with a platform prototype ensuring flexible, secure, robust, interoperable, pay-per-demand access to information, communication and processing infrastructure, see Fig. 1. This introduction will present some of the key findings and developments related to the semantic grid architecture for virtual organizations and primarily distributed engineering for the AEC industry. Additionally, the introduction part will present key user roles that have been identified in the platform as well as crucial end user requirements. This will be followed by a description of the system architecture including conceptual, service frameworks and key developed components that the platform offers. In this case, the following main '5S' challenges have been identified in InteliGrid:

- Security - industry eager to move only to a ground-up secure environment,
- Simplicity - must work seamlessly with current client applications and operating systems together with personalization and natural user interfaces,
- Standards - a need for stable long-term specifications,
- Service oriente architecture - scalable, adoptable, reuse, agile,
- Semantics - must support rich, domain specific semantics.

Recently, Cloud computing has been promoted as a solution to many challenging problems listed above. In fact, at both conceptual and technological levels, Cloud computing has many roots in Grid computing. Nevertheless, with much better support from various IT and independent software vendors, Cloud computing could today offer various capabilities needed by enterprises at the production rather than the prototype level as it was experienced back in InteliGrid. For instance, Cloud-based solutions provide dynamic or even event-driven configuration capabilities that can be easily reused and adopted to various changing conditions of business processes. Additionally, Cloud computing has introduced new system management technologies based on its core technology - virtualization by isolating different applications and operating systems from the underlying hardware. Automatically, virtualization technologies also solved many previously encountered security problems in Grids related to user-level access control or firewalls as the whole hosted operating system together with all user's applications and data are separated from the underlying core operating system mechanisms. Cloud computing could also help users scale their software in aggregate conditions. To conclude, it may be stated that Cloud computing wins by leveraging automation, virtualization, dynamic provision, massive scaling and multi-tenancy. Even though some new solutions have been proposed, the above-identified challenging problems are still present. Nevertheless, Cloud computing has introduced new challenges as advanced technologies development has led to power (mainly for cooling) and IT administration becoming dominant costs in data centers. Thus today, many researchers are working on resource management techniques to support energy efficiency in distributed virtualized IT systems [16][14][2].

### **3 The ACGT Project: e-Infrastructure for the Medical Research and e-Health Platforms**

Providing people with personalized treatment, post genomic clinical trials in particular, was one of the key challenges addressed by another EU-funded project, Advancing Clinico-Genomic Trials on Cancer (ACGT) (2006-2010)[20]. In this case, based on an advanced state-of-the-art analysis and obtained results, another two key challenges have been identified. The first important challenge in carrying out medical research as well as providing better ICT-based services for an active health monitoring and early diagnosis was an efficient access to and management of sensitive patient data located in many heterogeneous sources. In most cases, patient health records have already been collected in various hospitals, clinics or laboratories without the need for further data processing, semantic



**Fig. 2.** An overall distributed IT architecture of the ACGT platform with main actors and processes involved

annotation or interoperability. Moreover, patient records were protected behind many healthcare organizational barriers and shared under legal constraints that may vary significantly between countries. The second key challenge was related to an active participation of a patient in the whole process of secure data gathering, sharing and access control, which became extremely difficult with increasing cross-border patient mobility and the advent of ubiquitous Internet connections and new services.

Within the ACGT project we have successfully demonstrated an innovative approach which addresses aforementioned challenges by linking technical expertise in semantic and grid technologies together with legal, ethical and data-protection issues surrounding patient data. The solution has been defined in the form of a framework consisting of a combination of technical, organizational and legal measures. Even if the ACGT project focused mainly on clinico-genomic trials on cancer, our databases included a wide variety of anonymous data, such as: symptoms, histology, administered treatment, response to treatment, imaging data (X-ray, MR, US, CT, etc.), genomic data, or pathology data. To achieve syntactic integration, data access services first need to provide uniform and secure data access interfaces. This includes uniformity of transport protocol, message syntax, query language, and data format as well as data protection. Through the ACGT platform, in particular syntactic access services, data can be queried over the Internet using SPARQL, thus hiding the different query mechanisms provided by the underlying databases.

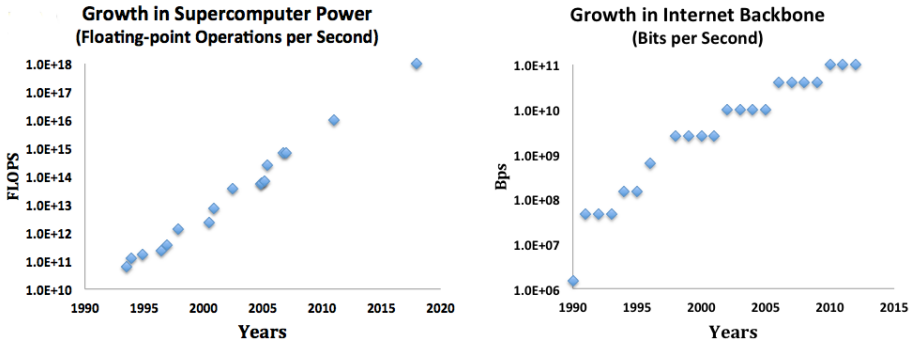
From the technological perspective, a set of interdependent and reusable software components, including Grid and Web services and access tools form the foundation of the entire ACGT platform, see Fig. 2. All security related components and message/transfer-level protection mechanisms are largely based on

a set of proven concepts, technologies and implementations [1]. Thus, risks associated with purely technological aspects have been limited. However, due to architecture and design limitations or simply because it is technically impossible to prevent all cyber attacks and actions, technical security measures leave a number of open gaps. Therefore, we identified the need for a separate governing body such as Centers for Data Protection in international collaborative environments for coordinated technical and procedural auditing, especially in the case of sustained operational and transnational eHealth services. Owing to innovative and powerful data integration and exploitation tools as well as multi-scale modeling and large-scale simulations that have been demonstrated in the ACGT project, we may also envision innovative healthcare services concerning and integrating all levels from the molecular and basic organs to the living organism that will also be a part of the virtual health records. Thus, not only basic historical data will be collected in the trusted networks, but also more sophisticated data structures integrated with personalized computing models for a better prediction and treatment of diseases will be available. Many best practices and lessons learned from this project could pave the way for new ICT approaches for future platforms, especially in the e-Health domain, fostering bio-medical knowledge acquisition and offer better health care services at the regional, national and international scale [3].

## 4 The CaKernel Project: A Tool for Emerging Complex Computing Architectures

According to the famous Moore's law, the number of transistors on a chip will double approximately every two years. In fact, capabilities of most digital electronic devices are strongly linked to Moore's law: processing speed, networking bandwidth, memory capacity, sensors or even a resolution of multimedia displays. All of these have been improving at roughly exponential rates and dramatically enhanced the impact of digital electronics in nearly every segment of the world economy. Processors vendors always tended to implement their own benchmarks by optimizing applications and programming models for their new IT systems delivered to the market. However, there are some well-known application benchmarking suits that were commonly accepted to evaluate the overall computing power over the last twenty years, in particular LINPACK [11]. Fig. 3 shows the exponential growth in Top500 supercomputing power measured by the LINPACK application benchmark and the networking bandwidth in Internet backbone.

One should understand that the computing power has increased by a factor of one million over the last forty years reaching the level of 10 000 000 000 000 FLOPS (10 Peta FLOPS) in 2011. For comparison, a hand-held calculator performs relatively few FLOPS, whereas majority of personal computers are capable of performing at the level of hundreds of Giga FLOPS. With the available 100GB/s optical networking connections, in fact using simultaneously multiple connections, users are able to move 1 petabyte from one place to another in



**Fig. 3.** The improvement at roughly exponential rates of capabilities offered by digital devices, for instance a) growth in supercomputing power, b) growth in Internet backbone

about few hours. One petabyte of data is roughly equivalent to few billion good quality digital photos and Google processes more than twenty petabytes of data per day [9].

The Moore's law can also be of relevance in the case of yet another computing paradigm shift. Many users have not noticed a huge shift behind the transition from single-core to multi-core CPUs in 2005. From smartphones, through desktops, up to huge supercomputers, processing units have been permanently transformed into heterogeneous computing clusters. All the changes in hardware design have also an enormous impact on the software stack, from operating systems to software languages, programming tools and finally applications. A single compute intensive application will need to harness the power of heterogeneous cores, and in practice it means for software developers that majority of applications have to be reimplemented from scratch based on new parallel data structures, algorithms and synchronization routines. Metaphorically speaking, the Moore's law hit the wall in 2005 and since then has kept many software developers busy with application redesigning, reimplementation, and performance testing. Moreover, to effectively exploit new capabilities provided by multi-core processors, their modern architectures and low level APIs, software developers must be aware of many factors that will impact the overall performance of their applications. From the hardware perspective, one should take into account not only the clock rate and a number of cores, but also check the memory bandwidth, efficiency of communication channels, cache topologies, etc. On the other hand, from the software perspective, it has become extremely important to understand data dependencies, data structures and synchronization among multiple parallel tasks for a given problem as all these factors will play a critical role during the execution [13].

With the recent advent of new heterogeneous computing architectures, there is still a lack of parallel problem solving environments that can help scientists to use existing computing facilities easily and efficiently, in particular new hybrid

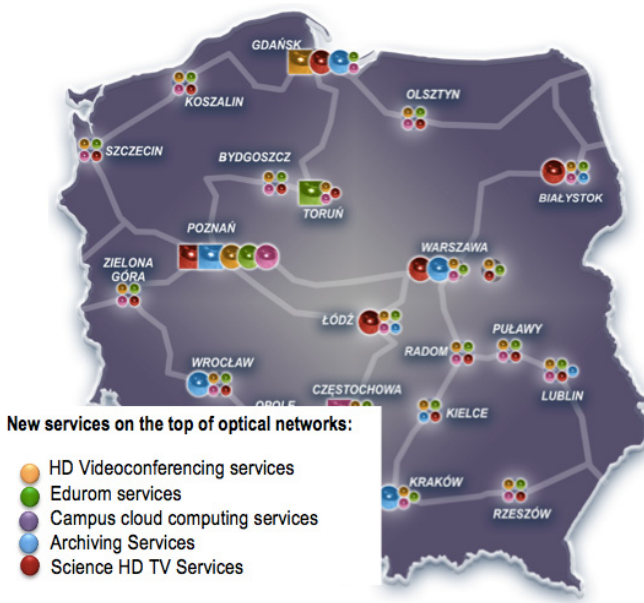
supercomputers. For instance, many scientific simulations which use structured grids to solve partial differential equations in fact rely on stencil computations. Stencil computations have become crucial in solving many challenging problems in various domains, e.g. engineering or physics. Although many parallel computing approaches have been proposed, in most cases they solve only particular problems. As a result, scientists are struggling when it comes to the subject of implementing a new stencil-based simulation, especially on high performance hybrid supercomputers. Therefore, in response to the presented need, an example of parallel programming framework for different parallel programming environments will be presented, in particular recently created the CaKernel framework [5]. In a nutshell, the framework is a tool that simplifies the development of parallel applications on emerging hybrid computing architectures. CaKernel has been built on the highly scalable and portable Cactus framework. In the CaKernel framework, Cactus manages the inter-process communication via MPI while CaKernel manages the code running on GPUs and interactions between them. As a non-trivial test case, the performance and scalability of the automatically generated code from CaKernel will be presented.

## 5 The Platon Project: e-Infrastructure for the e-Science

Nowadays, an advanced analysis of complex, large-scale, multidimensional data sets is recognized as a key component in many e-Science areas, including computational fluid dynamics, medical imaging, life science, and computational engineering. Modern supercomputers, as it was briefly discussed in the previous section, especially hybrid clusters based on accelerated graphics, have recently reached the level of petaflops and are able to easily produce petabytes of data. In the near future so called Big Data will be obtained from huge scientific instruments, such as satellites, telescopes, or large-scale sensor networks, accelerators, but also from a large number of commonly used instruments as their prices have dropped significantly, e.g. high-resolution CT medical scanners or gene sequencers. Additionally, recent developments in the fields of High Definition (HD), 4K, 8K or even 16K video capture and projection, rich multimedia devices, high-speed data networks and storage hardware, and advanced digital image compression algorithms, are making ultra high digital broadcasting, on demand visualization, video streaming and digital cinema feasible and in the future easily available for everyone [7]. Moreover, there are expected to be 1 trillion new mobile devices, with high-speed networking interfaces and high resolution panels, connected to the Internet in the near future, which will help drive even more exponential growth by the year 2020. One should note, that the exponential growth in Big Data poses significant scalability and performance challenges for traditional data analytics, social media monitoring and business intelligence solutions. Deriving knowledge from large data sets presents specific scaling problems due to a large number of items, dimensions, sources, users, and disparate user communities. A human ability to process visual information can augment the analysis, especially when analytic results are presented in real-time iterative



and interactive ways. What we are witnessing is another shift in computing generating information out of distributed data sets in real-time. There will be many e-Science communities that will benefit from this advancement.



**Fig. 4.** The successful deployment of five key e-Science PLATON services on the top of PIONIER high-speed optical network with the 100Gb/s backbone inside Poland and up to 40Gb/s cross-border connections to different countries

Over the last decade, the successful development of the national high-speed optical PIONIER network allowed researchers to design and deliver innovative e-Science services to support researchers in Poland [17]. Thus, the main goal of the follow up PLATON project (2009-2012) was to deliver a set of key ICT services [4] [19] [15]: videoconference, eduroam, campus clouds, archiving and science HD TV services to foster research and scientific collaboration within and outside Poland, see Fig. 4.

- Videoconference Services: realized by building a high-quality, secure videoconference system in the PIONIER network enabling point-to-point connections, as well as simultaneous connections among multiple locations in Poland, giving the possibility of recording and replaying particular videoconferences from remote data archiving services.
- Eduroam Services: easy to manage, secure roaming and single-sign-on solution for people from the scientific and academic communities in Poland by launching model secure systems of access to the wireless network in every MAN network and HPC centre.

- Campus Services: Cloud services offered on the top of an innovative computing infrastructure consisting of a large number of small clusters deployed on campuses which deliver applications on demand thanks to virtualization technologies, capable of providing a wide range of campus users, mostly students and researchers, an elastic and scalable access to specific applications, both in MS Windows and Linux systems.
- Archiving Services: available at a national level, offer remote archiving and backup capabilities as a value added to the national, academic and research PIONIER network. Archiving services, which increase the real-time data protection, are one of the elements necessary to increase reliability of functioning of each unit, and are targeted at academic environments, including the higher education system, research and development units and clinics dependent on universities and medical universities.
- Science HD TV Services: a set of a national level services offering interactive science HD television based on high definition digital content for both education and popularization of science and telemedicine.

## 6 Conclusions

Recently published the Cyberinfrastructure Framework for 21st Century Science and Engineering (CIF21) presents a set of key complementary and overlapping components that are critical to effectively address and solve the many complex problems facing science and society [22], [23]. Core components include data, software, campus bridging and cybersecurity, learning and workforce development, grand challenge communities, computational and data-enabled science and engineering, and scientific instruments. It is clear that e-Science is increasingly based on global collaborations that need access to different resources (e.g., people, instruments, libraries, data, computations, software, algorithms, etc.) across different science communities (e.g., Research Infrastructures, Grand Science Challenges, Societal Challenges) which are at different levels of maturity and of different sizes and impact (e.g., niche high-end science, campus science, citizen science, etc.). To deliver sustainable e-Infrastructure across all of these areas, it is necessary to engage with representative stakeholders from across all of these areas. It is vital to understand their usage scenarios and resulting requirements from which a multi-year plan that identifies and delivers the operation, maintenance and development of the key common components that can be agreed across the communities. A sustainable future for the digital science community that allows for maximal science return through easy controlled sharing between domains requires at the global scale solutions to the following challenging problems in the near future:

- easy access to e-Infrastructure providers and their core services at the national and international levels,
- the dynamic provisioning of high-performance networking,
- the services and tools needed to provision and use e-Infrastructure,

- people should not use directly Big Data, but information created on-demand via appropriate tools and services,
- the data services of curation, discovery and movement are needed.

Each of these areas will continue to evolve along different timelines and costs associated with them. To ensure long-term sustainability for what will be a multi-year plan where continuity and vision is essential, engagement with policy makers and funders.

The service-driven Information System domain becomes a new complex domain, which requires new interdisciplinary approaches. If and how all the identified and briefly discussed challenges for e-Science as well as lessons learned from the past research will influence specific business and enterprise processes are still open questions. Nevertheless, to deal with the increased complexity of inter-organizational and intra-organizational processes it is worth to be aware of innovative ICT solutions successfully implemented and deployed on e-Infrastructures to be able better design, implement, deploy and integrate your own IT solution.

## References

1. Adamski, M., Kulczewski, M., Kurowski, K., Nabrzyski, J., Hume, A.: Security and performance enhancements to OGSA-DAI for Grid data virtualization. *Concurrency and Computation: Practice and Experience* 19(16), 2171–2182 (2007)
2. Bak, S., Krystek, M., Kurowski, K., Oleksiak, A., Piatek, W., Weglarz, J.: GSSIM - a Tool for Distributed Computing Experiments. *Scientific Programming* 19(4), 231–251 (2011)
3. Binczewski, A., Kurowski, K., Mazurek, C., Stroinski, M.: A Concept of a patient-centered Healthcare system based on the virtualized networking and information infrastructure. In: *Proceedings of The Third International Conference on eHealth, Telemedicine, and Social Medicine*, pp. 51–58. IARIA (2011)
4. Binczewski, A., Starzak, S., Stroinski, M.: Academic MANs and PIONIER - Polish Road e-Infrastructure for e-Science. In: Davoli, F., Meyer, N., Pugliese, R., Zappatore, S. (eds.) *Remote Instrumentation Services on the e-Infrastructure*, pp. 193–208. Springer (2011)
5. Blazewicz, M., Brandt, S.R., Kierzynka, M., Kurowski, K., Ludwiczak, B., Tao, J., Weglarz, J.: CaKernelA parallel application programming framework for heterogeneous computing architectures. *Scientific Programming* 19(4), 185–197 (2011)
6. Castejon, F., Reynolds, J., Serrano, F., Valles, R., Tarancn, A., Velasco, J.: Fusion Plasma Simulation in the Interactive Grid. *Computing and Informatics* 27, 261–270 (2008)
7. Ciznicki, M., Kurowski, K., Plaza, A.: GPU Implementation of JPEG 2000 for Hyperspectral Image Compression. In: *SPIE Remote Sensing Europe, High Performance Computing in Remote Sensing Conference*, Prague, Czech Republic (2011)
8. Czajkowski, K., Fitzgerald, S., Foster, I., Kesselman, C.: Grid Information Services for Distributed Resource Sharing. In: *10th IEEE International Symposium on High Performance Distributed Computing*, pp. 181–184. IEEE Press, New York (2001)
9. Dean, J., Ghemawat, S.: MapReduce: simplified data processing on large clusters. *Communications of the ACM - 50th anniversary issue: 1958 - 2008* 51(1) (January 2008)

10. Dolenc, M., Klinc, R., Turk, Z., Katranuschkov, P., Kurowski, K.: Semantic Grid Platform in Support of Engineering Virtual Organisations. *Informatica* 32, 39–49 (2008)
11. Dongarra, J., Luszczek, P., Petitet, A.: The linpack benchmark: past, present and future. *Concurrency and Computation: Practice and Experience* 15(9), 803–820 (2003)
12. Kazi, A.S., Hannu, M., Laitnen, J.: ICT Support Requirements for Distributed Engineering in Construction. In: Stanford-Smith, B., Chiozza, E. (eds.) *E-work and E-commerce: Novel Solutions and Practices for a Global Network Economy*, pp. 909–915. IOS Press (2001)
13. Kopta, P., Kulczewski, M., Kurowski, K., Piontek, T., Gepner, P., Puchalski, M., Komasa, J.: Parallel application benchmarks and performance evaluation of the Intel Xeon 7500 family processors. *Procedia Computer Science* 4, 372–381 (2011)
14. Kurowski, K., Oleksiak, A., Piatek, W., Weglarz, J.: Hierarchical Scheduling Strategies for Parallel Tasks and Advance Reservations in Grids. *Journal of Scheduling* 11(14), 1–20 (2011), doi:10.1007/s10951-011-0254-9
15. Pekal, R., Stroinski, M.: PLATON - Advanced Service Platform for e-Science - Capabilities and Further Development Directions, Towards and Beyond Europe 2020 the significance of Future Internet for regional development Future Internet Event Report. In: *Proceedings of The Future Internet Conference, Poznan* (2011) (to appear)
16. Rozycki, R., Weglarz, J.: Power-aware scheduling of preemptable jobs on identical parallel processors to minimize makespan. *Annals of Operations Research* (2011), doi:10.1007/s10479-011-0957-5
17. Rychlewski, J., Weglarz, J., Starzak, S., Stroinski, M., Nakonieczny, M.: PIONIER: Polish Optical Internet. In: *The Proceedings of ISThmus 2000 Research and Development for the Information Society Conference, Poznan Poland*, pp. 19–28 (2000)
18. Seidel, E., Allen, G., Merzky, A., Nabrzyski, J.: GridLab - A Grid Application Toolkit and Testbed. *Future Generation Computer Systems* 18(8), 1143–1153 (2002)
19. Stroinski, M.: Innowacyjnosć PIONIERa szansa dla przyspieszenia realizacji celow strategii lizbonskiej w Polsce. In: *Proceedings of the 1st Conference i3: Internet - Infrastruktury - Innowacje*, pp. 7–21 (2009)
20. Tsiknakis, M., Brochhausen, M., Nabrzyski, J., Pucacki, J., Sfakianakis, S., Potamias, G., Desmedt, C., Kafetzopoulos, D.: A Semantic Grid Infrastructure Enabling Integrated Access and Analysis of Multilevel Biomedical Data in Support of Postgenomic Clinical Trials on Cancer. *IEEE Transactions on Information Technology in Biomedicine*, 205–221 (2008)
21. Worldwide LHC Computing Grid Technical Site,  
<http://lcg.web.cern.ch/LCG/Default.htm>
22. Cyberinfrastructure Framework for 21st Century Science and Engineering,  
[http://www.nsf.gov/about/budget/fy2012/pdf/40\\_fy2012.pdf](http://www.nsf.gov/about/budget/fy2012/pdf/40_fy2012.pdf)
23. Advanced Computing Infrastructure: Vision and Strategic Plan,  
<http://www.nsf.gov/pubs/2012/nsf12051/nsf12051.pdf>