

Information Systems to Support Collaboration in Large Research Projects

Lars Wolter, Wei Min Wang, and Rainer Stark

Abstract—Large scale research projects at research institutes and universities bring together multiple researchers from different disciplines, with different goals together. For successful collaboration between the project partners, such projects need to be supported with suitable information systems that allow them to do research within an environment with changing requirements and goals. Thus the information system needs to be able to adapt to any kind of developments during the course of the project in order to ensure efficient research activities. To successfully implement such systems in large research projects technical solutions and organizational best practices are needed. This paper presents solutions and best practices developed within the Collaborative Research Center 2016 – Sustainable Manufacturing.

Index Terms—Large scale research projects, collaboration platform, adaptable information systems.

I. INTRODUCTION

Research projects constitute a special kind of projects with unique characteristics. Since scientific research becomes more and more interdisciplinary national and international funding organizations are increasingly relying on multi-institutional research collaborations [1].

In Germany the German Research Foundation (DFG) funds so called Collaborative Research Centers (CRC). CRCs in general address complex topics that require interdisciplinary collaboration which is a common and useful practice even across multiple institutes [2]. Despite of a centralized organizational structure with a speaker and dedicated management board CRCs usually lead to high efforts in the development of efficient coordination and communication techniques. Due to the complex nature of such projects, where the results come from different disciplines with different wordings from different partners, it is also difficult to transfer results into the public in order to create a process of social feedback or to gain awareness. To meet those challenges the DFG embedded “Service Projects for Information Infrastructure” (INF) in CRCs which provide innovative IT-solutions to support collaborative research activities and the knowledge transfer to the public.

The goal of this paper is to present experiences made during the course of the CRC 1026 - Sustainable

Manufacturing (CRC 1026) from the perspective of an INF project in providing an information system to support collaboration in the CRC. Based upon these experiences collaboration barriers are derived and solutions and best practices are presented. Additionally, the paper is also colored by the experiences the writers gained across multiple projects done at the Technische Universität Berlin and the Fraunhofer Institute, publicly founded as well as contracted by industry.

A. CRC 1026 - Sustainable Manufacturing

The CRC 1026 is an interdisciplinary research project funded by the DFG. It started in 2012 and consists of nineteen subprojects involving the four disciplinary science clusters of manufacturing, environmental engineering, economics and mathematics. The scientific theme is “From saturated markets, bridging the gap to hungry markets”. In fact, more than 60 researchers will collaboratively explore ways to face the economic, environmental and social challenges of future development by exploiting potentials of sustainable manufacturing and embedding them in to global value creation. The subprojects, each addressing one specific aspect of sustainable manufacturing, are organized in three major project areas and three cross-sectional subprojects.

Subject of Project Area A - Strategy Development is the identification and development of pathways for sustainable technology, their assessment, valuation and mathematical modeling.

In Project Area B - Production Technology Solutions manufacturing processes and equipment, virtual systems for product development and organization of sustainable value creation in product and material cycles on different levels of aggregation will be realized exemplarily.

Project Area C - Principles, Methods and Tools for Qualification links these perspectives by developing tools and methods enabling humans for learning and teaching help for self-help.

Cross-sectional projects include Public Awareness, research training group and the service project for information infrastructure.

B. Service Project for Information Infrastructure (INF)

The INF-Project is involved in administration tasks and extends the software tools the CRC needs for collaboration. In the beginning a common web platform for researchers, project partners and the public audience was established. Aside of basic features like email distribution lists, shared data repositories and discussion boards, innovative collaboration processes, methods and tools are developed and integrated into a collaboration platform build upon a Java based web application called Liferay [3]. Social media

Manuscript received January 1, 2014; revised March 3, 2014. This work was supported by the Deutsche Forschungsgemeinschaft (SFB 1026/1 2012).

Lars Wolter and Wei Min Wang are with the Technische Universität Berlin, Industrial Information Technology, Berlin, Germany (e-mail: lars.wolter@tu-berlin.de and wei.min.wang@tu-berlin.de).

Rainer Stark is with the Fraunhofer Institute for Production Systems and Design Technology, Division Virtual Product Creation, Berlin, Germany (e-mail: rainer.stark@ipk.fraunhofer.de)

features are applied to enable and facilitate involvement of all subproject partners.

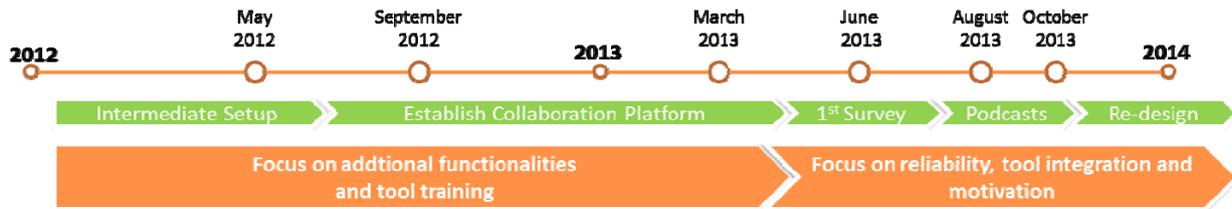


Fig. 1. Time line of INF activities and focus areas since start of the CRC.

The goal of the INF project is to ensure long term accessibility of research results of the CRC 1026 and to establish a knowledge platform for sustainability and related topics.

II. EXPERIENCES GATHERED DURING THE COURSE OF THE INF PROJECT

The special characteristics of CRC-like projects affect multiple areas, not only the information systems but also organization structures and processes of all kinds of tasks. Research projects in research institutes and universities have often fluctuations in personnel due to project oriented working contracts and work for dissertations. CRC-like projects have multiple partners acting in different disciplines and in different institutes.

Until now the INF project went through different phases (see Fig. 1) in the beginning of the project it was necessary to focus on establishing basic functionalities to support the researchers. These functionalities included shared repository, contact management and a website for the public audience. This preliminary setup was implemented using a Microsoft SharePoint service hosted by the central IT of the TU Berlin. Parallel the development of the final platform was initiated. After extended testing the platform was deployed with basic functionalities. The release step was accompanied by training sessions for the researchers. Until 2013 the platform was extended by functionalities from the preliminary solution. Once the platform was in constant use the focus switched to reliability, tool integration and motivation to raise the usage of the platform. The start of this phase was driven by a survey to retrieve the current requirements and IT competences of the researchers. Based on this survey the functionalities in the platform have been redesigned and extended. The relaunch was accompanied by podcasts and blogs to keep the users informed and involved. During this phase a taskforce of key users was also implemented. This phase is still ongoing. Throughout the phases the INF project experienced various obstacles that can be categorized into issues regarding requirements management and motivation of the users.

A. Requirements Management

Requirements management is the first big hurdle during software development projects. Users are often not able to formulate their needs resulting in large costs and broken deadlines. Research projects like the CRC with their interdisciplinary collaboration can also raise new requirements during the work phase of the project that could not been thought of beforehand. This also means that some of

the defined requirements will never be required during the project even if they are specified. Because of the different origins of the project partners implicit requirements need to be considered, that originate from single requirements. For example the need that the information system are to be embedded into tools used by the project partners can implicitly define the requirement of platform independence if one project partner is using Linux and another one is using Windows operating systems. But none of those persons will require platform independency on them self.

Another specialty common to research projects like the CRC are the changing persons during project runtime. The most common change is between the application and start of the project, which can take more than a year. Therefore the requirements formulated during the application by one group of persons are not the requirements of the persons working on the project later on.

B. Motivation Problems using New Information Systems

Researchers tend to work in their own small enclave doing most of their exchange with other researchers using tools like email or thru papers and presentations. For the day to day research work they use their own tools and infrastructure which is not defined by a higher authority. An example would be the usage of word processors, one uses Latex another uses Word and they exchange their knowledge by using read-only PDF. During the CRC those different researchers need to work together more closely requiring the use of common tools to be efficient. Because there is no common toolset or even method or process to do specific tasks some of the researchers need to use new tools aside from their own tools which results in less motivated usage.

Because we incorporated a web-based platform during our project we could monitor the usage of that platform showing low usage except for situations where the users where forced to use the platform, before specific deadlines for example (see Fig. 2).

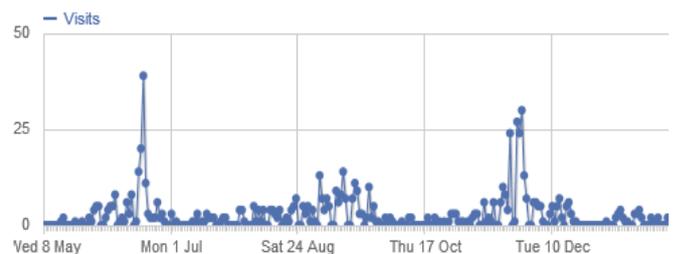


Fig. 2. Usage statistics of the CRC platform. The first peak was the first overall project meeting, where every subproject needed to supply a status to the platform, the peak at the end was the second project meeting and the middle cluster of smaller peaks was a conference dedicated to the CRC.

III. SOLUTIONS TO ADDRESS THE REQUIREMENTS MANAGEMENT

Because the information infrastructure needs to be present at the very beginning of the project, it is very important to know the basic needs of the collaboration partners before start of the project. Therefore, a recording of basic and abstract requirements needs to be done before starting the project, as soon as the start has been accepted. This verifies the most important requirements from the application of the project and matches them with the current setup of available and commonly used tools and methods across the project. This recording of requirements should also check the aims of each subproject to allow an infrastructure project to foresee possible upcoming requirements or possibilities to support the research activities.

After starting the project, a user centric survey needs to be done to gather the current requirements and IT competences of the users. The time for this survey should be at a point where all personal is established and all the subprojects that need to collaborate have at least talked once with each other. With this knowledge the specific requirements of single users can be considered. This includes their usage of methods and tools as well as the type of media they use (phone, e-mail, social networks). The survey also allows gathering the needs for collaboration. Is there only document exchange necessary or do far away positioned partners need web conferencing and concurrent editing solutions.

During the runtime of the project the personal responsible for the information systems should establish a feedback channel using different media to allow the communication of additional requirements or changes of requirements.

Nevertheless additional surveys need to be done in regular intervals to ensure that any changes in the requirements are known as soon as possible even if they are not communicated directly by the researchers. An example of a survey result showing the prioritization of collaboration features is presented in Fig. 3.

Rank	Feature	Average ranking
1	Document versioning	1,86
2	Workflow management	3,09
3	Shared Calendar	3,14
4	Management dashboard	3,59
5	Polls	4,18
6	Yellow pages	5,14

Fig. 3. Priorization of coordination features. during the survey the participants needed to order the different features for their personal importance. (1=important,6=unimportant)

IV. SOLUTIONS TO ADDRESS THE MOTIVATION PROBLEMS

Addressing the problem that the users of the CRC where not motivated enough to use the platform resulted in different solutions. Some of these solutions where successful, but others were not. Important when trying to raise the motivation to use a specific IT-system or platform is that the target users really need this platform to be more efficient in their research. Motivating the usage of something nobody currently requires might drive researchers into a situation of wasting time which limits their research activities. Therefore a requirement for this kind of motivation is that the users are

motivated to collaborate with each other.

A. Involve the Users while Evolving the System

As with usual software development practices in smaller projects where agile methods [4] are commonly used, it is important to be as near to the customer as possible. The customer in case of the CRC are the researchers and therefore it is important to constantly engage with them and presenting them new possibilities and features to ensure that they will address their requirements as expected. This way they are motivated to using something they are part of during development and not only using something someone else dictated them to use. It proved very effective to establish a group of voluntary key users that regularly meet up to involve the users instead of trying to involve everybody.

The implementation of new features was communicated using different media like blog entries, webcasts or showing them different scenarios they usually do side by side with the use of our platform and without. This did not seem to generate the desired effect of involving the researchers. For them to get an advantage from the implemented system and all its features they need to get in contact with them during their usual work.

B. Integrate New Tools with the Existing Toolset

Bringing the look and feel of existing tools to the new platform and integrating it into the regular working tools makes the tools much easier to use. A lot of development has been done to connect every feature with the corresponding office tools to allow a more fluid working for a large number of users. The integration can happen on multiple levels and can begin with using single-sign-on mechanisms based on current authentication systems, making calendars accessible through outlook or allow the office tools to directly edit content on the platform. In case of the CRC-1026 it was easy to implement the read only export of the calendar to outlook as well as the direct editing using WebDAV. More complicated was the reception of outlook meeting emails to automatically insert them into the online calendar. Single-Sign-on is still not implemented because of security restrictions of the existing authentication mechanisms

C. Using Social Media

The usage of social media to raise the interest in a platform is a well-established course now-a-days. It includes the possibility for all users to share their knowledge by posting information, writing comments on or rating other postings. The working of social media is based on experiences of the users with public social media platforms. But not everybody is using these platforms and is therefore used to their usage patterns. During one of the surveys the CRC-members were asked how much they use different applications and platforms in their other activities, resulting in very low numbers regarding social media (Fig. 4).

This survey showed why the CRC was not reacting very much to the social media features incorporated and gives hints, why surveys that gather the user experiences and requirements are so important and that there are many facts to consider when implementing social media [5].

Another important factor that determines the applicability of social media mechanisms is the critical mass of users.

Usually, only five to ten percent of social network users contribute actively while the rest remain passive consumers. Such network members who share knowledge, experience and opinion actively are the core drivers of social networks. They create valuable contents that attract other users to join the network or encourage passive users to participate in a guided way (e.g. by starting and moderating discussions). Thus, a critical mass of users is crucial to the success of social media mechanisms. In terms of the CRC the number of users is quite limited. Taking into account the fluctuation of personnel and temporarily involved guest researchers, there are about 80 people on average who can be considered as permanent CRC members, including principal investigators, researchers and student assistants. As the usage statistics (see Fig. 4) implicates this is not enough to create a critical mass for the successful application of social media mechanisms.

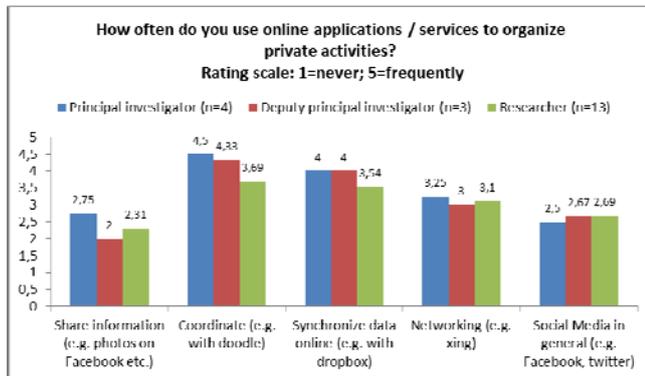


Fig. 4. Survey results showing the usage of various types of tools the users use during their private activities. The survey distinguishes between the different user roles to accommodate for the processes these roles are involved with.

D. Using "Big Data" Methods

During the last years there has been a large increase in possibilities to raise the size and quality of information that can be retrieved from structured and unstructured data. The extraction of data from various data sources and their processing as well as the establishment of links between the processed data is called big data, if this data is too big to handle, too slow to handle or too complicated to handle by using usual data analytics [6]. So it might depend on the actual project how big the data really is. Giving the users of an information system the possibility to retrieve new information related to their project by applying big data methods can raise the motivation to use the information system. For the big data processing to be effective many data sources are required, so individual users can put information about their research into the information system and get back possible links to other research the colleagues have done and have not been expected. Such results give the possibility to establish new fields of collaboration with new partners in the CRC.

The motivation can be raised even more if outside data sources are incorporated. This includes for example map data or open statistics data from the government. But if these external data can be useful and the effort to integrate them is feasible depends on the aims of the subprojects.

For the CRC-1026 a search mechanism was implemented that could search across all the different media types available on the platform. External data sources have not yet

been implemented but the integration of sustainability databases is ongoing.

E. Using Ontologies

In addition to extracting information from various data sources it is important to have consistent knowledge models that are developing through out the project. A possibility to map these models is to use ontologies. The ontology developed for the CRC-1026 allows the subprojects to connect their knowledge, even if there are differently used terms and definitions. The ontology allows the researchers to work on a common understanding of their project gathering information about possible connections with other subprojects and to better use synergies between them. An information system which incorporates the ontology gives a motivation for them using the information system with the feeling that it understands their own view of the project [7].

The ontology build in the CRC-1026 connects different disciplines like production, virtual engineering, mathematics and economics. The ontology features many terms with multiple usages across all the disciplines giving the researchers important hints when they collaborate with different disciplines. To further use the ontology a deeper incorporation into the information system is planned to aid the search tools by adding possible links through the ontology.

V. BEST PRACTICES AND RECOMMENDATIONS TO ESTABLISH AN INFORMATION SYSTEM FOR AN CRC

During the first two years of the CRC-1026 the INF-project gathered a lot of information regarding successful tasks and less successful tasks to effectively support the whole CRC. Some best practices can be formulated on these findings as well as based on literature in this area. The best practices can be differentiated between the technological level, the organizational level and the disciplinary level. The technological level handles the implementation of IT-Systems, technical features and methods regarding the technical solution while the organization part focuses on processes and soft skills. The disciplinary part considers the abstract level of the research subject.

A. On an Technological Level

Selecting the appropriate IT-Systems to build up the information system architecture for CRC-like research projects involves other challenges besides the requirements described above. One of the important decisions is between a monolithic solution and heterogeneous IT-landscape. The monolithic approach has its advantages in the reduced integration efforts but has flaws in the details of specific features. Using a heterogeneous IT-landscape allows the use of best-in-class-systems to address the most important requirements and choose simple solutions for seldom required features, but generating a lot of work integrating all the different systems. For the CRC-1026 a monolithic solution was chosen because it contained all the features necessary to address all requirements defined in the application of the project. The solution build upon a Java

based web application called Liferay. Liferay is a CMS and platform to support collaboration and can easily be extended by writing so called portlets. The platform delivered a lot of features ready to go that needed only marginal updates in design and configuration. But after working with this solution some of the requirements changed and new ones were defined. Three important aspects could be examined after starting with the monolithic approach and trying to adapt it to the changing project. Those three favor a heterogenic landscape:

- 1) Only a few of the functionalities of the monolithic system are useful, because the others are still not used by the researchers. One of these important functionalities used was the document management which was constantly used from the first day of and was not best-in-class in terms of usability and stability. Features like portal-wide chat or the usage of personal blogs have never been used even if they were required during the project application.
- 2) Not all parts of the infrastructure need to be connected and therefore no integration effort is needed for those parts. In the CRC-1026 case this happened for the external website which does not need any connection to the intranet area because only one subproject manages this site and all content is exclusively in this part of the system.
- 3) The most important fact supporting a heterogeneous IT-landscape is the evolvement of the monolithic system to a heterogeneous landscape. Because of the changing and new defined requirements during project runtime additional features needed to be implemented, that the monolithic solution could not or not easily supply. Two examples for this are the Web-Conferencing-System which needed to be integrated into the UI and user management of the monolithic solution and the additional Wiki-Solution that should allow for an independent user management to be used by open public (see Fig. 5).

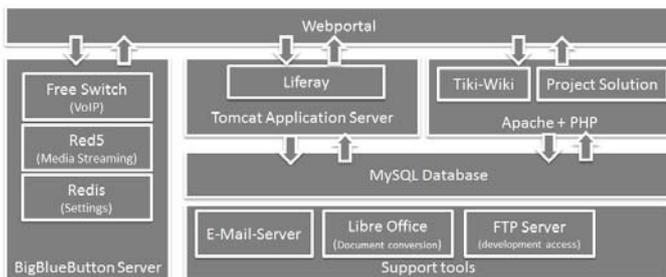


Fig. 5. IT Infrastructure of the CRC-1026 information systems.

Besides those three facts the use of a monolithic solution is still viable if the requirements can be expected to be stable. It is also more advantageous in case of maintenance. The servers to manage for the CRC now include two machines to accommodate all the needed features and also a third machine for testing purposes. If changing requirements and collaboration needs are expected, a heterogeneous landscape of specialized solutions is more advantageous in supporting the researchers.

B. On an Organizational Level

Something common to interdisciplinary projects across multiple institutions is the reduced outcome of the projects because of higher coordination efforts [8]. The high coordination effort is partially caused by the lack of consistent processes in the beginning of the project. The experiences during the INF project showed that it is not only necessary to establish a good working information infrastructure but also the business processes that use this infrastructure. The initial top-down processes that have been defined in the beginning of the project have been found faulty because they did not interact well with discipline specific organizational structures. This resulted in situations where these processes were not followed when executing tasks. The assertion of those processes is complicated because all of the participants are more or less equitable and the management board is missing the right of command and therefore only executing coordinative tasks. Establishing a decision-making body failed consecutively because the commitment of the subprojects was too vague.

This might occur because of the additional workload involved with those decision-making bodies. But one of the more important points is, that each subproject has its own interests that it needs to focus on and which are partially in competition with other subprojects or do not allow to satisfy the needs for the overall project results and collaborative desire. Business- and working processes should therefore be defined and enforced by a representative decision-making body right from the beginning. This management board needs all the authority to give directives. The infrastructure subproject can help in cooperatively define those needed processes showing potentials and borders how processes and IT-infrastructure can aid the collaborative research.

A problem also occurs because of two different types of collaboration. In the first case specific, very determined and highly formatted work deliveries are needed, like overall project reviews. In the other case there are informal and pro-active collaborations without formal boundaries. The distinction of these two cases is not always clear for the researchers. This ambiguity leads to situations where researchers don't share unfinished results because of the fear they might not be correct in terms of formal boundaries. The processes and the information system need to clearly state which kind of collaboration is expected from the researchers by declaring message boards to informal areas for example.

C. On an Epistemic Level

The maturity level of the epistemic domain of CRCs should also be a determinant for the information infrastructure. As Corley *et al.*, [1] suggest the success of research collaboration also depends on the maturity level of the scientific discipline. During the development process of a scientific discipline from a burgeoning research topic to a full-fledged scientific discipline the collaboration mode and partners will most likely change. Therefore, the requirements for information infrastructure may also vary from level to level. A research group which works on a burgeoning research topic would probably need more support in terms of coordination and finding common baselines while working on a full-fledged scientific discipline would require tools for

professional exchange, publication management and archiving of information and research results.

The collaborative research topic of the CRC, sustainability, per se is yet not a full-fledged scientific discipline as it is still used in diverse contexts and with different meanings. This became also clear as multiple discussions were lead between the project partners about the definition of sustainability during the first phase of the CRC. Another challenge within the CRC was to find proper interfaces between the partners. To address these challenges the ontology based graphical glossary mentioned above is planned to visualize connections between disciplinary terms and subjects and to help all partners to identify interdependencies and collaboration potentials.

VI. TRANSFERRING THE ASPECTS GAINED FROM THE CRC-PROJECT TO OTHER PROJECTS

One of the main differences between CRC-like research projects and projects in the industry is the defined environment. There are projects that might be structured in a similar manner as the CRC, that run for specific period of time and include the collaboration of different departments of a single or even multiple companies. But the difference is that in a single company the processes, methods and information infrastructure are more centralized than in the CRC, but can be loose, too. Even in projects across multiple companies there is always a lead which has a well-established environment and integrates those partners on a technological and organizational level.

There are multiple collaboration problems in the industry [9] especially concerning the motivation problems, making the solutions and practices presented here applicable to projects in the industry. This is also the case for research projects to other larger projects where information systems need to be established for collaborating partners.

VII. CONCLUSION

The paper presents the experiences gained while implementing and managing the information infrastructure of a Collaboration Research Center project and offers solutions and best practices commonly encountered during the start and runtime of such large research projects. The experiences during a large CRC-Project when applying solutions like social media, customer involvement and big data for example are described to give the reader a good understanding as well as recommendations and best practices.

The best critical practices described are

- 1) Acquire relevant and precise requirements before and during the project.
- 2) Supply IT solutions to the critical working activities like document management.

- 3) Use customization abilities of open source solutions to integrate existing toolsets and working behaviors.
- 4) Propose and promote methods and processes for collaboration.
- 5) Close involvement of the customer by implementing advisory boards.

It is also shown how some of these solutions can be transferred to different kinds of projects.

REFERENCES

- [1] C. A. Elizabeth, B. P. Craig, and B. Barry, "Design and the management of multi-institutional research collaborations: theoretical implications from two case studies," *Research Policy*, 2006, vol. 35, no. 7, pp. 975-993.
- [2] T. A. Finholt, "Collaboratories," *Annual Review of Information Science and Technology*, 2002, vol. 36, no. 1, pp. 73-107.
- [3] J. X. Yuan, *Liferay Portal 6 Enterprise Intranets: Build and Maintain Impressive Corporate Intranets with Liferay*, Packt Publishing, 2010.
- [4] P. Abrahamsson, O. Salo, J. Ronkainen, and J. Warsta, *Agile Software Development Methods: Review and Analysis*, VTT Publications 478, 2002.
- [5] A. M. Kaplan and M. Haenlein, "Users of the world, unite! The challenges and opportunities of social media," *Business Horizons*, vol. 53, no. 1, pp. 59-68, 2010.
- [6] S. Madden, "From databases to big data," *Internet Computing, IEEE*, vol. 16, no. 3, pp. 4-6, 2012.
- [7] W. M. Wang, A. Pfortner, K. Lindow, H. Hayka, and R. Stark, "Using Ontology to Support Scientific Interdisciplinary Collaboration within Joint Sustainability Research Projects," in *Proc. 11th Global Conference on Sustainable Manufacturing*, 2013, pp. 612-617
- [8] J. N. Cummings and S. Kiesler, "Collaborative research across disciplinary and organizational boundaries," *Social Studies of Science*, vol. 35, no. 5, pp. 703-722, 2005.
- [9] L. Wolter, H. Hayka, and R. Stark, "Improving the usability of collaboration methods and technologies in engineering," *IFIP AICT*, vol. 411, pp. 73, 2013.



Lars Wolter was born in Berlin, Germany and graduated in computer science at the Freie Universität Berlin. Since then he works as the chair of Industrial Information Technology at the Technische Universität Berlin, focusing his research in the area of product lifecycle management and collaboration in engineering



Wei Min Wang was born in Shanghai, VR China and graduated in Business Engineering to the Technische Universität Berlin. He is a research assistant to the chair of Industrial Information Technology at the TU Berlin since 2012. His research focuses on scientific and engineering collaboration and ontology application for knowledge management



Rainer Stark studied mechanical engineering at Ruhr University Bochum and Texas A&M University. From 1989 to 1994 he worked as research assistant at the chair of Design Engineering / CAD of the technical faculty of the University of Saarland. His PhD thesis is entitled "Development of a mathematical model on tolerance for the integration in (3D-) CAD systems". From 1994 to 2008 he worked at the Ford Motor Company in the area of virtual product creation. Since February 2008, Rainer Stark has been a professor for Industrial Information Technology at the Technical University of Berlin and the director of the Virtual Product Creation Division of the fraunhofer institute for production systems and design technology (IPK).