A Life Cycle Model for Collaborative Research Environments

Stefan Buddenbohm[†], Harry Enke[‡], Matthias Hofmann[*]
Jochen Klar[‡], Heike Neuroth[†], Uwe Schwiegelshohn[*]

Abstract: Virtual or Collaborative Research Environments are important pillars of modern research in many scientific disciplines. Therefore, the development and particularly the sustained operation of these environments has become an important challenge in the area of modern research infrastructures. In order to support communities in dealing with this challenge including design and evaluation, we have developed a life cycle model for Collaborative Research Environments. Based on this life cycle model, we identify common pitfalls and suggest a generic catalog of criteria to be used by developers, operators, and funding agencies. To consider the heterogeneity of Collaborative Research Environments, additional discipline specific criteria can be incorporated.

1 Introduction

Across many disciplines modern science often requires large amounts of digital research data and various software products. Therefore, information technology has become vital for most research environments. Contrary to expectations published in many magazines the use of IT in research has led to a significant increase of expenses during the last years, which are not compensated by technological progress. But IT also offers new opportunities as it enables us to access the components of our research environments from most locations of the world due to the availability of high bandwidth connectivity almost everywhere. This location-independent access allows us to share the components of our research environment with many colleagues throughout the world and thus leads to new and more efficient research collaborations. We can also distribute the costs of maintaining large data centres or specialized research software across many different interest groups. These synergy effects offer the potential of at least partially compensating the rising costs of science. As such research environments are shared by research groups belonging to different research institutions, it is often called a virtual research environment to emphasize that the cooperation of people from different institutions forms a *virtual organization*. Unfortunately, this this name is misleading as the word 'virtual' is often regarded as the opposite of 'real'. But any virtual research environment is real for both the user and the provider of

[*] {matthias.hofmann, uwe.schwiegelshohn}@udo.edu, Robotics Research Institute, TU Dortmund University
[†] {buddenbohm, neuroth}@sub.uni-goettingen.de, Göttingen State and University Library
[‡] {henke, jklar}@aip.de, Leibniz Institute for Astrophysics Potsdam (AIP)

the environment. Since the mentioned connectivity also supports communication between different research groups working on the same topic, we prefer the expression *collaborative research environment (CRE)* although collaboration does not comprise all uses of the research environment.

As the potential benefits of collaborative research environments are recognized in many disciplines, a significant number of such research environments has been created recently. But we must realize that despite the interest in collaborative research environments and the growing number of new collaborative research environments, sustained operation of a collaborative research environment is rarely achieved. In a project supported by the German research council, we have have examined collaborative research environments in various research disciplines and related IT infrastructures. This study has led us to the characterization of different phases in the life cycle of a collaborative research environment. We are able to specify different pitfalls in these phases. There are various reasons for these obstacles: sometimes the effort to transfer a prototype environment into a production environment is underestimated, sometimes legal and administrative procedures hamper establishing an environment that is shared by different institutions in different states or countries. We strongly believe that only a thorough understanding of the complete life cycle of a research environment will help to overcome these obstacles. Therefore, we will explain this life cycle and the different aspects of its phases and connected success criteria in this paper.

The paper is organized as follows. While Section 2 deals with related work, Section 3 describes the methodology that has been applied. In Section 4, we suggest a life cycle model for CREs, and describe its characteristics. The paper concludes in Section 5.

2 Related Work

Our work follows up on numerous studies which have been recently performed in this field, whereby we focus for obvious reasons on the discussion in Germany. First and foremost the work group "Virtual Research Environments" of the Priority Initiative Digital Information by the Alliance of German Science Organisations[dAIDI11] is of importance since we use their discussions and studies as preliminary work for our approach, especially their concept to assess the degree of maturity or establishment of CREs within their respective communities. Based on the work of the 'Joint Information Systems Committee' (JISC) in the course of the 'British Virtual Research Environment Programme'[CR10], it became clear to us that we must choose a pragmatic approach emphasizing the demands of researchers and developers/ operators.

An alternative approach could have focused on the funders' perspective and the evaluative aspects of measuring research infrastructures, e.g. to facilitate funding decisions, like the ERINA, and ERINA+[1]-projects funded by the European Commission within the FP7-Programme. Decisive criteria for ERINA+ were the satisfaction of the user community and the amount of CRE-assignable results.

[1] http://www.erinaplus.eu/

The lack of comparability of the results remains an unsolved problem for all these approaches. Although it is possible to evaluate the quality of CREs based on reproducible facts, it is in our eyes impossible (or at least questionable) to compare these evaluations.

Furthermore, the work group 'Virtual Research Environments' of the German Initiative for Network Information (DINI)[For11], the DFG project 'Radieschen'[For13b], and the WissGrid project of the German D-Grid Initiative[LE13] were important for our work as they contributed valuable experiences from various past and current projects.

Our modular CRE life cycle entailed fruitful discussions in various interviews and workshops (see Section 3). Although some generic aspects may also fit into software engineering models, CRE specific inspiration was specifically derived from the work of the 'Cooperative University Computing Facilities' (SURF) with its VRE Starters Kits[SUR12].

3 Methodology

Ideally a collaborative research environment supports the complete research process or at least major parts of it. Since CREs are relevant to many different scientific communities and there are significant differences in the research process between different communities including tools and methodological approaches, we cannot expect that a single scheme for a CRE life cycle model and success criteria fits all communities. But due to the efforts in time and money required to establish a successful CRE, we also cannot afford to build every CRE from scratch. Instead, we must identify common properties and processes within CREs to allow guidance and evaluation how to generate and operate a CRE. This is in particular of interest for operators of CRE and funding agencies.

Similar to large software systems there are different approaches to build and run a CRE. Instead of using a theoretical approach, we decided on collecting information from existing CREs or related systems by conducting expert interviews with suitable stakeholders based on our life cycle model (LCM). In addition to institutions that operate CREs, also e-Infrastructures and software providers such as European Grid Infrastructure (EGI), the German National Research and Education Network (DFN), and the Higher Education Information System (HIS) provided valuable input as they face problems similar to CRE, especially regarding financial issues. The interviews were conducted over a period of three months. Since our interview partners provided expertise in different areas, we tailored the interviews individually instead of using a fixed scheme. As some interviews have yielded results that were surprising to us, we decided to incorporate such results into the questions of later interviews to determine whether these results are important in general or are only relevant for a particular CRE.

Additional input was collected within two workshops. The first workshop reviewed the results of the interviews. To this end, all interview partners and additional representatives from different scientific communities, CRE operators, and funding agencies were invited. The second workshop was organized by the German Research Council and discussed organizational and management issues for virtual research environments.

As a result of our study, we have developed a life cycle model that is suitable for most

CREs. This LCM comprises five phases (see Section 4) and was considered to be very useful by the participants of the first workshop where it was presented for the first time. We structure the main part of this paper according to the phases of the model. However, it is noteworthy that the establishment of a CRE might not strictly follow the very life cycle, i.e. iterations between individual phases like in current software development models occur. Thus, the life cycle model describes an idealized view that aims to guide through the establishment and operation of a CRE.

Regarding the evaluation of a CRE or - to be less controversial - to determine best practices for their generation and operation of a CRE, we propose to define a generic set of core criteria and milestones. Key objective is the usefulness of the CRE to researchers and its efficiency in operation. Due to the heterogeneity of CREs, we suggest that the generic success criteria serve as a pool from which an individual set must be selected in each case depending on the scope and service portfolio of the specific CRE. Furthermore, it is necessary to supplement these generic criteria with discipline-specific or community aspects and criteria since, for instance, a CRE in the arts and humanities at least partially requires different criteria than a CRE in engineering or natural sciences.

4 Life Cycle of a Collaborative Research Environment

The life of a CRE can be divided into five different phases. The schematic sequence is displayed in Fig. 1 while a list of success criteria is given in Table 1. In the first phase, the *prototype*, a first version of the CRE is constructed and deployed for a limited number of users. This phase can demonstrate the overall feasibility of the system and provide first assessment of the acceptance by the corresponding scientific community. If the prototype is, as we will discuss later, considered successful, the subsequent phase of *development* is used to address all aspects, which are crucial for a sustainable version of the CRE, but were left out while developing the prototype. An operational concept is most important in this context, the services of the CRE are matched with the required resources and a budget in which the financial resources and the corresponding funding sources are specified.

Again, if a set of criteria and milestones is met, the CRE enters the next phase. For a more infrastructure-like CRE we call this phase *operation*, while for exclusive software projects the term *product* is better suited. In this phase the CRE is made available to the corresponding scientific community. During this time, a special focus lies on the support of the users of the CRE and the subsequent technical and organizational improvement. When, after some time of operation, the decision to discontinue the CRE is made, a phase of *transfer* is used to move the components of the CRE which are worth conserving (e.g. research data, source code, etc.) to other (infra-)structures. The remaining technical and organizational components of the CRE are then dissolved in the phase of *liquidation*.

In general, there are two different ways of operating a CRE. Either the CRE is built and operated by a collaboration as a separate infrastructure unit. This case is mainly considered in the following. Or the CRE is embedded as a module in a broader infrastructural context, e.g. as a part of a data center. Here, prototype and development phases, as well as transfer

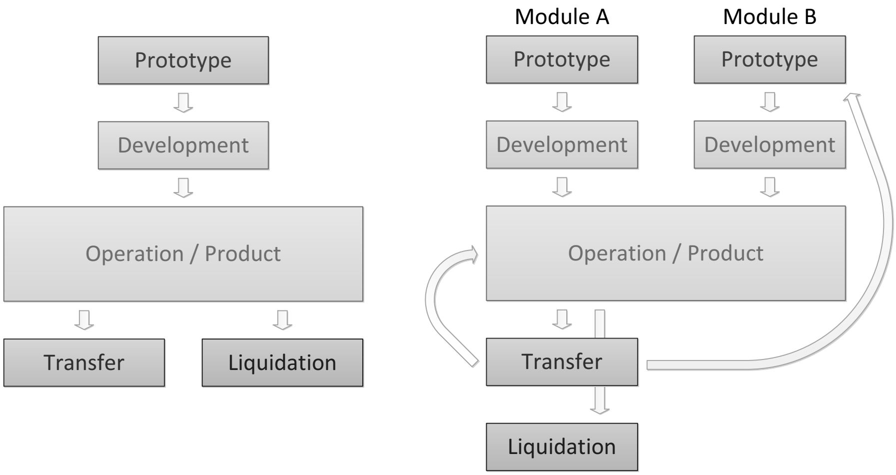

Figure 1: The life cycle of a monolithic CRE is shown in the left figure. After the creation of a prototype and a subsequent development phase, the phase of operation (infrastructure) or product (software) is reached. When the CRE is discontinued, components of the CRE which are worth retaining are transferred, while the rest is liquidated. The right figure shows the complex life cycle of a modular CRE. A first module A is created similar to the first scenario and passes the prototype and the development phases. Then it enters a operational phase in a shared infrastructure with other modules. At the end of its lifetime some of its components are transferred to modules still in the phase of operation, while others are used in the creation of a new module B. The remain of module A is dissolved in the phase of liquidation.

and liquidation follow the same rules, but are applied to this specific module only. As shown in Fig. 1, parts of a CRE can also be transferred not only to external infrastructures but may be reused in all stages of a new CRE module.

4.1 The Prototype

The first step towards the implementation of a new CRE is the creation of a prototype. In general this is carried out in form of a research project including the usual review process. Accordingly, one of the first steps (usually part of the initial application) is to specify the internal organizational structure of the project and how the work is distributed. The first steps of the project itself then comprises a set of *preliminary studies* concerning the environment of the new CRE. Those studies include a detailed *demand analysis* as well as a *determination of the relation* of the new CRE within the context of existing infrastructures, both locally (e.g. computing centres, libraries) and with respect to the main research community of the CRE (e.g. data centres). Moreover, a *positioning* with respect to the other stakeholders in the research community of the CRE, a *thorough analysis* of scientific work flows which the CRE aims to support, and a *market analysis* of the considered hardware and software components are needed.

We consider these preliminary studies to be success criteria in a broader sense; they are essential for progressing in the further development of the CRE. Building upon these analyses, the further process is split into two branches, a technical branch in which the technical infrastructure is set up and the software development takes place, and an organizational branch, which is responsible for the development of corresponding organizational structures. In our model, these branches are further divided into specification and implementation.

During the *technical specification*, the services of the CRE are defined based on the preceding analysis of the scientific work flows. The software is being developed using a well-known, established process which is usually the waterfall model. Moreover, standards and as metadata are chosen, and decisions for the used hardware and software are made. In the subsequent *technical implementation*, the software development work is carried out, and the installation of hardware and software takes place. Extensive testing of the system (i.e. unit, integration tests) ensure a running system.

The *organizational specification* comprises a round-up of the different access permissions for the resources of the CRE (e.g. different layers of access to research data) as well as possible policies. In addition, strategies for integration to the community and the creation of support facilities such as a help desk and training group need to be prepared. The *organizational implementation* phase is then used to raise awareness in the corresponding research field and build a community around the new CRE. The set-up of preliminary support facilities and first training sessions for future power-user can already take place at this point.

The described works culminate in a *pilot system* in which the new prototype is made available to a limited number of scientific users. This is not only used for further testing and refinement of the system, but also offers the possibility to adjust the CRE through smaller improvements to meet the needs of its community.

In order to gain information on the particular technical, financial and organizational requirements, especially with respect to a sustainable modus of operation, all described phases are accompanied by a systematic monitoring, which is part of the project management.

4.2 Development Phase

During the pilot phase of the prototype, first impressions of the working system and in particular an assessment of the user acceptance can be done. Based on this work, a decision for the proceeding of the next phase of the CRE, the development phase, is made.

With respect to financing CREs, it is beneficial to distinguish between the first two phases (i.e. *prototype* and *development phase*) that build up or extend the very product (CRE), and the operation phase. While the build up phases are often financially supported in form of projects, the operation phase requires a sophisticated business model to ensure potentially long-term provisioning of services. Therefore, it is of vital importance to integrate a business plan into the planning efforts at this early stage.

The phase of development is then used to prepare all necessary features the CRE needs for a subsequent sustainable phase of operation, but which are generally not implemented while creating the prototype. This includes the creation of a *legal entity* for the CRE and a *concept of operations* in which the provided services to the community are related to the required resources. Also a financial budget needs to be prepared. Another important task is the compilation of a sustainability concept, which specifies the whereabouts of different components (e.g. research data, software) of the CRE after the end of its lifetime.

4.3 Operation

Besides the technical operation of the CRE, key aspects in this phase are the support of the community and the continuous improvement of the technical and the organizational environment. While no strict temporal sequence of tasks is given, a set of equally important tasks can be identified. An overview of these tasks is given in Fig. 2. In general the critical relevance of user involvement and community acceptance for the success of the collaborative research environment cannot be underestimated. The presented life cycle phases and success criteria consider this aspect by clearly defined tasks (monitoring, periodical evaluation, feedback from users), by preliminary studies before the start of the development phase and later on by dedicated criteria that allow an evaluation of user involvement and community acceptance.

Support for the users is provided by a help desk and a sufficient documentation of the CRE. Through focused events (e.g. workshops, training) the community of the CRE is maintained and expanded. Public relations advertise the CRE in the scientific community and to a wider audience. Although, the CRE should provide a stable working environment for the users, it is necessary to perform further work on the system. Especially requests from users and new features emanating from new user groups require substantial development. Also security issues of the used software components and following updates need to be taken into account. The coordination of these aspects lies within the requirements management. While larger improvements are organized as separate projects that might follow a similar life cycle model as the CRE, smaller improvements and bug fixes can be performed within the structures of the CRE. In addition to the production system, a CRE should have a development and a testing system to integrate developments seamlessly into operation. In conclusion the criteria for the success of a CRE during the operational phase can be summarized as following:

Community-related aspects, e.g. the acceptance of the CRE within the potential community of users, the degree to which the CRE meets the specific demands of the users or its ability to improve its services alongside a moderated process of communication between users and developers of the CRE.

Performance-related aspects, e.g. the ease of use (not just usability) of the CRE, its preferably seamless integration into already existing research environments of the users and its stability and operational availability.

Result-related aspects, e.g. the exploitation of synergies within the research work of the

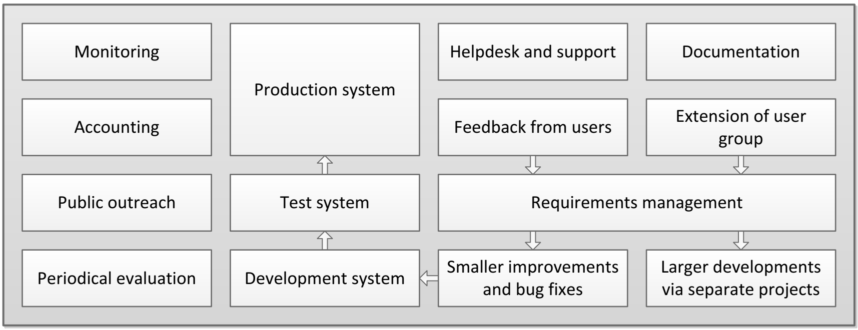

Figure 2: Central aspects of the operation (or product) phase of a CRE.

users. Based on our interviews during the course of the project the synergies are at the forefront of many users' expectations towards a CRE and not the enabling of new research questions with the help of technology. The result-related aspects emphasize the character of the CRE as a research tool for its users, as a means to ease their daily work. Furthermore aspects of collaboration and the use of distributed resources are to be subsumed here.

Transfer-related aspects, e.g. the occurrence and use of initial resources, experiences services, and products in other CREs or research environments.

Competence-related aspects, e.g. acquisition and transfer of individual and institutional expertise. These can be represented by staff as well as by documented experiences of infrastructure operators and developers that facilitate the development of subsequent services and products. It is also relevant how much a CRE affects the research practice of the respective discipline.

Of course the aforementioned remarks only represent a broad outline of the possible success criteria for a CRE. Each of the five *aspects* must be differentiated into detailed criteria, which are - ideally - as meaningful as possible. For each individual CRE a set of specific criteria must be selected out of the above mentioned aspects. Case-by-case these more generic criteria must be supplemented by discipline-specific criteria depending on the scope and services of the CRE.

Sustainable operation of CREs means that services should run as long as they are required by scientific communities as an integral part of their research. Regarding costs and financing, we clearly identify expenses for scientific and technical staff to be most relevant as a result of the expert interviews.

4.4 Transfer and Liquidation

At one point in time the CRE, or a smaller part of a modular CRE infrastructure, reaches the end of its lifetime, and the decision of its dissolution is made. In this case, the most im-

Phase	Criterion
Prototype	Demand analysis
	Technological landscape study
	Technologies / Tools / Standards
	Monitoring
	Community acceptance
Development phase	Concept of operations
	Organizational form
	Legal entity
	Financial budget
Operation / Product	User acceptance and satisfaction
	Enabling collaboration
	Support / Community
	Publications / Results
	Build-up of Expertise
	Public outreach
	Recurring evaluations
Transfer	Knowledge transfer
	Technology re-use

Table 1: Generic success criteria for CREs along the life cycle.

portant task is to transfer any components which are worth preserving to other infrastructures. These infrastructures can be external like community data centres or other CREs, or, in case of a modular CRE, different modules of the same system (see Fig. 1). Components which should be transferred are *research data* (which should be published or migrated to other data centres or CREs), *services* (which should be migrated to other providers), *software* (which should be published open-source or at least handed over to partners who can support them in the future), and *different other deliverables* (which should be published).

A transfer of these components can be mandatory, for instance, due to legal obligations or other commitments like the rules for safeguarding good scientific practice of the German Research Foundation (DFG)[For13a], but there can also be other reasons, like a still existing user community.

During the liquidation, the organizational structures of the CRE are dissolved. If required a successor organization can be founded. A migration support should help the last remaining users with the transition to different infrastructures.

5 Conclusion and Outlook

In this paper we described a Life Cycle Model for Collaborative Research Environments. The model is based on experiences and research done in various Grid- and other collaborative projects. It was refined by interviews and discussion with developers, operators, and users of CREs. In addition we drew on experiences and models of large software systems. Our interviews show that for CREs most efforts are currently spent on the prototyping phase, although experiences from development of large software systems show, that success of a product depends to a very large extent on early incorporation of constraints from operation and product transfer.

For the successful development of such a CRE it is crucial to consider all phases of the life cycle. This encompasses not only the developers and operators of such CRE, but also the funding partners or institutions, which have significant influence on the CRE. If the sponsors of CREs neglect to take the whole life cycle into account, the developers cannot be expected to resolve all connected issues by themselves.

We hope that our approach of combining general criteria and discipline specific elements into a life cycle model organizing the structural, functional, organizational, and financial aspects provides guidelines for all participants of CRE. These guidelines may help to build and operate successful CREs.

References

[CR10] Amy Carusi and Torsten Reimer. Virtual Research Environment Collaborative Landscape Study. http://www.jisc.ac.uk/publications/reports/2010/vrelandscapestudy.aspx, 2010. Accessed April 15, 2014.

[dAIDI11] Arbeitsgruppe Virtuelle Forschungsumgebungen der Allianz-Initiative Digitale Information. Virtuellen Forschungsumgebungen - Ein Leitfaden. http://www.allianzinitiative.de/fileadmin/user_upload/Leitfaden_VRE_de.pdf, 2011. Accessed April 15, 2014.

[For11] DINI AG Virtuelle Forschungsumgebungen. Workshop: Virtuelle Forschungsumgebungen - erste Erfahrungen und Ergebnisse. http://www.dini.de/veranstaltungen/workshops/forschungsumgebungen-2011/, 2011. Accessed April 15, 2014.

[For13a] Deutsche Forschungsgemeinschaft. Empfehlungen der Kommission "Selbstkontrolle in der Wissenschaft" - Vorschläge zur Sicherung guter wissenschaftlicher Praxis. http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/empfehlung_wiss_praxis_1310.pdf, 2013. Accessed April 15, 2014.

[For13b] Deutsche Forschungsgemeinschaft. RADIESCHEN - Rahmenbedingungen einer disziplinübergreifenden Forschungsdateninfrastruktur. Organisation und Struktur. http://dx.doi.org/10.2312/RADIESCHEN_005, 2013. Accessed April 15, 2014.

[LE13] Jens Ludwig and Harry Enke. Leitfaden zum Forschungsdaten-Management: Handreichungen aus dem WissGrid-Projekt. http://www.wissgrid.de/publikationen/Leitfaden_Data-Management-WissGrid.pdf, 2013. Accessed April 15, 2014.

[SUR12] SURFfoundation. Virtual Research Environments Starters Kit. http://www.surf.nl/en/knowledge-and-innovation/knowledge-base/2011/vre-virtual-research-eenvironment-starters-kit.html, 2012. Accessed April 15, 2014.