Neuantrag auf Sachbeihilfe

SPLICE—An experimental approach towards a web-native publication format

Proposal short name: SPLICE
October 27th, 2014

Beschreibung des Vorhabens – Projektanträge im Bereich „Wissenschaftliche Literaturversorgungs- und Informationssysteme“ (LIS)

LIS-Förderprogramm oder Ausschreibung:

Open Access Transformation

Aleksander Benjak
Academic co-worker (Akademischer Mitarbeiter) in Felix Wieland Group
Biochemistry Center of the Heidelberg University, Heidelberg

Contents:

1 Ausgangslage und eigene Vorarbeiten (State of the art and preliminary work)

Similar to the impact printing press had on creation of paper-based publications and journals three and a half centuries ago, web-based technologies are now providing the foundation for the creation of web-native knowledge dissemination formats. However, it is not only the way how scientists present their results that is changing, but the transformation of the entire scientific process is under way. Often, this movement is called “open science movement”.

Open science advocates are criticizing many aspects of the contemporary scientific process. They advocate against the long and inefficient publishing process, resulting in a large publication lag (the time from initial submission of a research article to its final acceptance for publication in a journal often takes many months). Impact factor which is currently used as a proxy for quality of research is deemed inadequate[1]. Peer-review is being debated because it is inefficient and struggles to cope with the demands of modern publishing[2]. The hyper-competitive culture in the scientific community that fosters quick publishing of only positive results at the expense of quality of the results, which very often cannot be reproduced, is criticized as well[3].

Changes to the current system are inevitable and many stakeholders (scientists, publishers, politicians and others) are playing a role in this transition. Most proponents of the open science movement will agree on some key features of the new scientific process. For instance, they believe all research results should be published in a digital form and made freely accessible to the research community and the public. They argue peer review process should be continuous and open, perhaps even crowd-sourced, with a new metrics and reputation system developed to measure both the impact and the quality of scientific research. Data, including negative results, should be shared in a standardized format that would allow computational processing of information. However, the differences arise between the different stakeholders in the perception of how the new system should look and how it should be implemented.

The current research cycle is complex, starting with a single observation or idea and formation of a hypothesis and ending with publication and review of research outcomes. It is therefore not surprising that the transition from a paper-based to a web-based publication will be complex as well. The typical view many stakeholders have is to narrowly focus on one particular feature. Examples include Figshare, a platform for sharing of experimental figures[4]; Altmetric and Impactstory, projects trying to develop metrics alternative to impact factor[5,6]; Rubriq, a project providing journal-independent peer review[7]; F1000, a publishing platform also experimenting with new forms of peer review[8]. This approach is facilitated by the availability of funding to support development, provided by venture capitals and publishers who are trying to exploit the new web-based landscape of scholarly publishing.

This kind of isolated evolution of the different components of the scientific process has led to some unwanted side-effects. As an example, open-access to publication, which is seen by most stakeholders as absolutely required, has been introduced into the system without proper assessment of risks. Open-access offers a new business model many aspiring publishers have jumped on trying to generate quick profits. The resulting inflation in the number of scientific journals and publications now seems too large for the available body of peers to cope with in the current peer-review quality control system. The end effect is sloppiness in publishing by some journals. This was demonstrated in a study published in 2013 which investigated open-access journals (although similar peer-review deficiencies are likely to occur in traditional journals as well)[9].

However, if open review had been implemented simultaneously with open access, these side-effects might have been avoided. The reason why such review system was not implemented at the time is because it was a much harder task. A study from 1999 showed that, although the majority of scientists were in favor of one form of open review, there was a significant difference in refusals to review when the same scientists were asked to review manuscripts anonymously (current system) and non-anonymously (new open review system). Non-anonymous reviews were more readily declined[10].

The example above illustrates how hasty implementation of individual changes can quickly influence the entire scientific system, because parts of the system are all interconnected and interdependent. As the open science movement strives to change the system as a whole, we believe a holistic approach towards development and implementation of a new scientific knowledge management and dissemination system might be a better solution. Such an experimental approach would provide valuable insights into the inter-dependencies in the system, which could guide development of prototypes and later production-quality tools for scientists. At the same time important information could be provided for policy makers about the implementation strategy and necessary interventions to make the transition towards open science possible.

Our goal is to build a prototype of a novel web-native publication platform, with integrated data sharing and reviewing features, and a community-based reputation system. Building such a platform that addresses many of the open science target features simultaneously will help us anticipate potential friction points, eliminate them and bring us closer to the goal of an integrated and well-functioning scientific workflow. In addition, based on usage data and feedback, we aim to create a set of guidelines, **a road-map towards open science**, which should help decision makers in adopting policies to facilitate this transition.

Previous work

In an effort to improve sharing and access to information, as well as increase efficiency of scientific research, I helped set up the non-profit organization Life Science Network gGmbH. This organization has been established in 2011 in Heidelberg, Germany, and remains entirely funded by contributions from the founders. As a volunteer, I played a significant role in the development of two web-platforms, which contain features relevant to this project.

The goal of the **Life Science Network** (http://www.lifescience.net) project was to develop a platform for professional networking and sharing of knowledge in life sciences. As a central feature, the platform contains a directory of life science infrastructure designed as a hierarchical tree, resembling the administrative structures in research institutions and universities. On top of this basic structure, various modules have been added, each supporting a particular type of content or activity. “Protocols”, for instance, supports sharing of experimental research protocols and recipes. The “Publications” module enables uploading of research publications. In total, there are eight different modules developed up to date.

The benefit of having an organized directory of infrastructure is that different kinds of content can be associated very precisely with the elements of the structure (e.g. institutes, departments, research groups) and, as a consequence, the relevant information is easier to be filtered and discovered.

As part of the Life Science Network project, we have also introduced an open review module. This is a system through which any scientist can submit his opinion on published research in several formats. Along with a commenting and recommendations system, numerical ratings as well as non-anonymous reviews have been implemented as well. Written review in combination with numerical ratings has already been described as a convenient format for open review[11].

Our approach was different than that of the typical altmetrics, best represented by Altmetric[5], a project promoted by Macmillan Publishers Limited (Holtzbrinck Publishing Group). Altmetrics approach consists of looking at different sources to collect usage or mention data about a given publication. Although impact factor has been heavily criticized by the altmetrics community and scientists in general[1], these novel metrics are similar to impact factor in the sense that neither provide the context in which scientific information has been used or mentioned. Online communities might heavily discuss (tweet, read, download, share) a publication due to its relevance, but also possibly because it is controversial, unreproducible or obviously fraudulent. Therefore, both metrics can be utilized as popularity indicators, but not as proxy for quality of research. If we wish to have a good metric for quality of research, we need new tools and new types of data that could provide the context in which information has been used. These tools might be similar to those developed as part of the Life Science Network project.

In addition, we have introduced a scoring system for activity of researchers, which allows us to dissect and precisely measure their contributions. Even more, due to the hierarchical nature of the structure behind the platform, it is possible to calculate the scores at the level of research infrastructure, such as research groups, departments or institutions.

The second project, called **Science Simplified** (http://www.scimplified.com) consists of a science communication web platform. The goal of this online platform is to bring news about scientific breakthroughs closer to the general public and to facilitate communication between scientists and the public.

Science Simplified features an easy way to upload news articles. In addition, it has a real simple syndication (RSS) feed aggregation in the background that integrates news from a large number of news sources (institutional RSS feeds). As the articles are imported, they are also automatically classified into different categories.

The RSS feeds and manually-uploaded articles can be associated with organizations, which are part of a hierarchical network, a tree-like structure similar to the one developed for the purposes of the Life Science Network project. The end result of this structure is that the user can finely define what kind of news he would prefer to get in his own personal news stream. The user can follow different topics, various authors or different places (not only organizations, but also cities and countries). The platform also integrates some typical social network features, such as commenting, recommendations, personal favorites lists and notifications, which are sent upon activity in the platform.

We believe this kind of communication between scientists and the public should be encouraged. Therefore, Science Simplified also features a scoring system that calculates contribution not only of individual authors, but also aggregated scores for institutions. In other words, we are able to quantify public outreach of research institutions and rank them by this score. We believe that such a scoring system should play a significant role when allocating public funds in the future.

Benefits to the project

There are multiple features developed as part of the Life Science Network and Science Simplified platforms which are of direct relevance to the SPLICE project.

First, detailed mapping of the scientific infrastructure will be important with regard to sharing and filtering of content, as it would allow authors of the content to share it with particular groups of people during preparation. Multiparameter ratings and open review features will be another aspect which could be implemented with minor modifications. Finally, calculation of scores for individual objects, and their aggregation at the level of researcher, or even groups, departments and institutes will be possible. The formulas for calculation of these scores will be made public.

SPLICE is envisioned as a partnership with the Life Science Network gGmbH. The project will benefit from the existing code, which has been written for the purpose of the Life Science Network and the Science Simplified projects, and will be contributed to the SPLICE project by the Life Science Network gGmbH. Using the expertise we gained during development of these projects, we would be able to shortcut development and launch the beta version of the SPLICE platform much earlier.

The project will also benefit from our experience in lean and agile web development, which we collected over the last three years. Life Science Network gGmbH has created and launched two projects with features beyond state-of-the-art that are on par with other platforms that received significant financial support from investors. We achieved competitor status with a manyfold smaller budget by careful planning and investment of time in the early development phase, in order to reduce the high coding expenses. On the other hand, expensive human-driven marketing has been replaced with an automated Google Adwords campaign (a grant by Google). These examples demonstrate Life Science Network gGmbH has acquired a unique skill set which will help us make the most out of the budget requested for this project.

1.1 Projektbezogene Publikationen (Project-related publications)

I have published no project-relevant publications. However, I played a major role in setting up the two scientific platforms relevant to this project: Life Science Network (http://www.lifescience.net) and Science Simplified (http://www.scimplified.com).

1.1.1 Veröffentlichte Arbeiten aus Publikationsorganen mit wissenschaftlicher Qualitätssicherung, Buchveröffentlichungen sowie bereits zur Veröffentlichung angenommene, aber noch nicht veröffentlichte Arbeiten (Peer-reviewed articles)

None.

1.1.2 Andere Veröffentlichungen (Other articles)

None.

1.1.3 Patente (Patents)

1.1.3.1 Angemeldet

None.

1.1.3.2 Erteilt

None.

2 Ziele und Arbeitsprogramm (Objectives and work schedule)

2.1 Voraussichtliche Gesamtdauer des Projekts (Project duration)

The project is intended to last 36 months with support of this grant. The project should continue beyond the funded period, but the format and scale of it will depend on the outcome and feedback of the research community and other stakeholders.

2.2 Ziele (Objectives)

Scientific research results are presently published in a paper-based form. But the system is under pressure by critics who argue that knowledge dissemination in such format is non-transparent, slow and inefficient. Open science movement is under way attempting to implement changes to the current system in several areas including: open access, open review, new quality and impact metrics etc. These approaches deal with individual features of a future open science system.

We propose a different, holistic kind of approach. Our main objective is to build a novel infrastructure, which will allow publishing research results in a web-native format, with integrated reviewing and metrics features. This platform will provide the basis for an integrated research workflow, and help us create a road-map towards open science implementation.

The SPLICE project will therefore directly address several of the objectives given in the program guidelines, including: set-up of an information system; optimal creation, unrestricted provision and dissemination of digital publications; modeling technical and organizational innovation in the area of electronic publishing; improving the acceptance of electronic publications and the open access paradigm; developing new forms of quality assurance for electronic publications; developing of tools for creating and jointly editing electronic publications; and merging articles, data and metadata.

O1: Building a prototype of a web-native publication platform for research results

The first objective of our proposal is to build a prototype of a web-native publication platform. The platform will also include reviewing features, a reputation system and a number of other features (see Table 1 below).

At the center of the current paper-based publication system is the article describing the hypothesis or idea, providing the rationale for the experimental approach, describing experiments and discussing results and their impact. Typically, more than one experiment is described in a traditional publication – a set of experiments is the minimal publication unit. The key assumption of our proposal is that the **elementary publishing unit should become smaller** than a typical paper-based publication. It should consist of a single bit of information, a single assertion (in size comparable to a figure in a traditional paper publication). We call these publication units **data objects**.

The benefits of reducing the size of the minimal publishing unit are described in Figure 1. and include increasing the speed of research, shortening the time of publication, reducing the duplication of effort and thereby cutting overall research expenses.

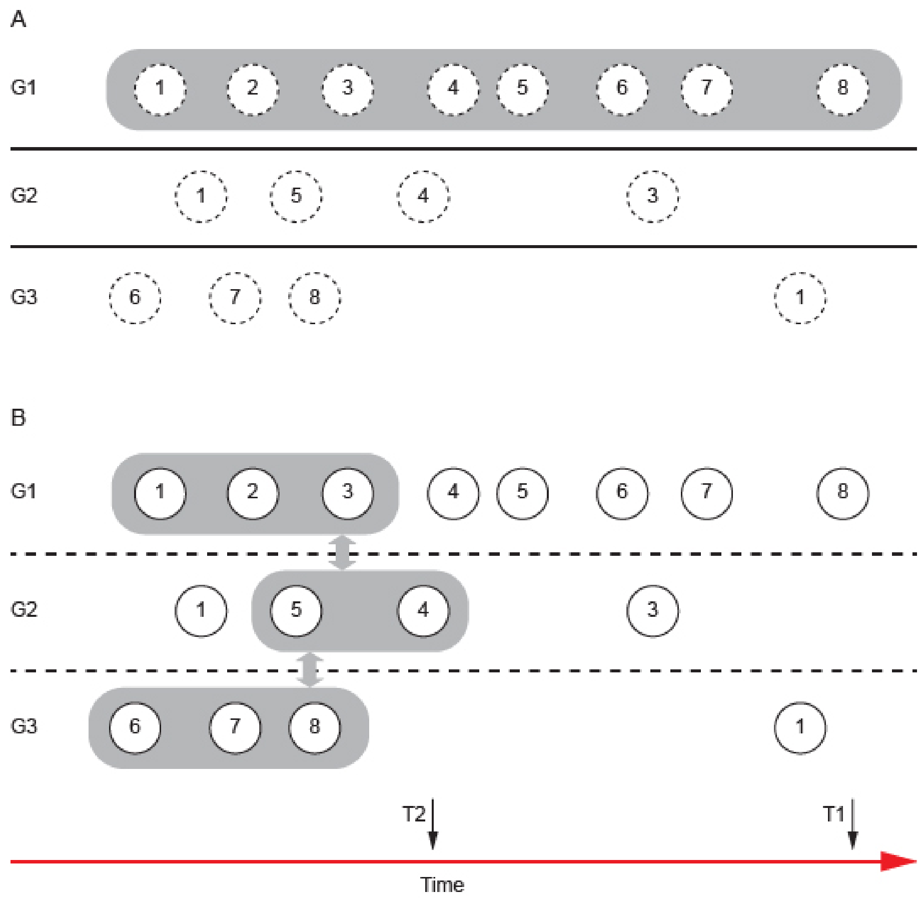

Figure 1. A scheme comparing the current system to the one described in our proposal. Numbered circles represent small pieces of information or individual assertions.
A) Typically, different research groups (G1-3 = research group 1-3) will work on the same scientific problem. They will collect and create bits of information and eventually connect and describe them in form of a narrative article and publish. Because there is no communication about the progress of research and sharing of data between the different research groups, even if other research groups (G2 and G3) obtain knowledge that could help solve the problem faster, this information if not available to the group number 1. The group 1 will eventually publish their results at time T1 indicated at the bottom of the figure. Groups 2 and 3 will be scooped.
B) If individual information packages (data objects) are published as soon as they are created, the information is instantly available to all research groups working on the same scientific problem. A set of data objects described with a narrative article (equivalent to the common paper-based publication) is now created much quicker than in the case above (at time T2), as it utilizes results from different research groups. Since data objects are cited, all groups are acknowledged for their contribution.

In their nature, data objects are very similar to the concept of nanopublication, which has been introduced recently[12]. However, in contrast to nanopublication, a data object should contain both machine-readable and human-readable content, and include a detailed set of structured metadata.

While there are other projects testing the idea of sharing smaller bits of information and metadata, including Figshare[4] and Scientific Data project (featuring Data descriptors)[13], our goal is to create a more complete system that would also support creation and publication of **data object sets**, collections of data objects, put together in a logical sequence and elaborated by a narrative article.

Filtering of data objects which can be combined into sets will be facilitated through the use of structured metadata. Only those data objects which match to a certain degree in their metadata can be combined into sets. Because there can be a large number of “classes” of metadata and this data is used to create relations between the individual data objects, we often refer to the structured metadata as to **dimensions of relation**.

The project name SPLICE is given because, similar to gene splicing, the construction of data object sets (and narrative articles) will have to occur by joining together individual data objects, all while respecting dimensions of relation. This concept is further explained in Figure 2.

The possibility to characterize data objects with metadata in a structured fashion, ensuring easy definition, search, filtering, interoperability and a mechanism for their integration into sets, will be a critical component of the project. The challenge will be to address the inherent complexity of the system, but create an acceptably intuitive user interface for scientists using it. Therefore, apart from including basic information, such as author(s), date, place etc., metadata will contain key details describing the experimental results according to field-specific standards of data presentations as recommended by the Minimum Information for Biological and Biomedical Investigations (MIBBI) Project[14].

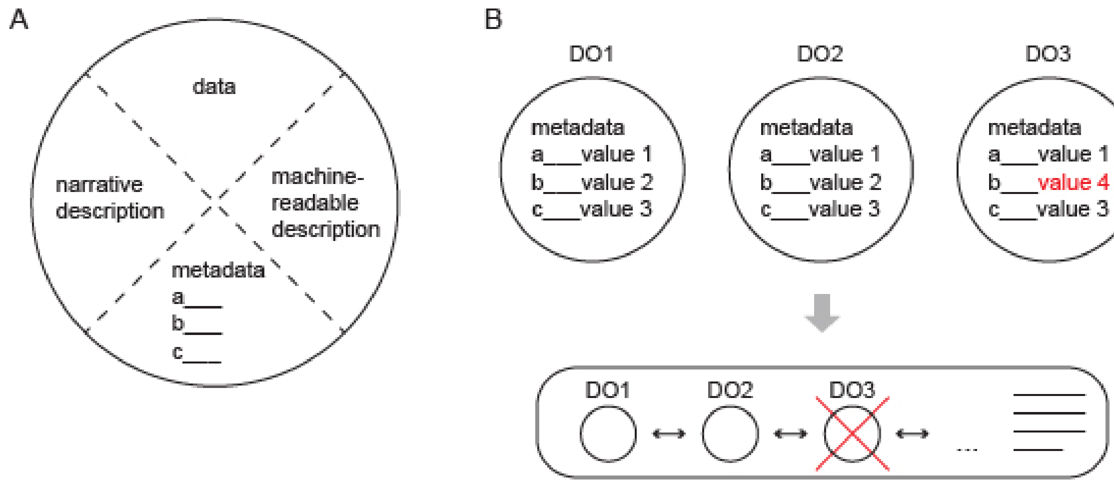

Figure 2. A scheme explaining data objects and structured metadata.
A) Each data object contains data (or information how data can be accessed), a narrative, human-readable description, a machine-readable description and a set of defined metadata parameters. The data object will have a time stamp and a unique digital identifier and handle so it can be cited.
B) When writing narrative articles, authors will choose from a pool of data objects they produced (or other scientists produced) and combine them into sets. Combining data objects with a higher overlap in metadata (here data objects DO1 and DO2) will result in more convincing and trustworthy stories and scientific conclusion.

At the level of data object sets and narrative articles describing them, the goal will be to create a system that allows multiple authors to edit the data object set, both during its creation and after publication. Since data object sets will be reviewed, post-publication editing of data object sets will enable scientists to address reviewers' concerns and comments, within an integrated version tracking system that will preserve the history of the revision process.

Another important feature of the system will be open review. We envision a system of written reviews in combination with numerical ratings as a central element of the system. This approach is in line with a study demonstrating that a combination of written reviews and numerical ratings is best suited for open review.[11]

A reputation system will be developed as well. It will be based on the activity of scientists (the amount of content they create) and the quality of the content (estimated by the community feedback and review scores). Development of this feature will benefit strongly from existing reputation systems developed by Life Science Network gGmbH for our previous projects (www.lifescience.net and www.scimplified.com). In essence, contribution will be built from points assigned for each data object, data objects sets or reviews. These points will then be further increased or decreased depending on the feedback of the community.

Additionally, we plan to integrate social networking features into the platform. For instance, the ability to follow and receive automatic notifications upon activity of certain users will be provided. Scientist will have a profile where they can include a summary about their research. A direct messaging system will be integrated as well. The list of the most important features is given in Table 1.

Feature	Description
• data objects	• elementary publishing units • contain data, metadata, narrative and machine-readable description • submission forms • search and filtering tools
• data object sets (narrative articles)	• submission forms • multiple authors • editable with version tracking
• reviews	• community-based • submission forms • numerical ratings
• researcher profiles	• information about the researcher
• institutional profiles	• information about the organization
• a tree of infrastructure	• information about the internal organization of institutions • local (branch-specific) search
• reputation system	• personal • institutional
• networking tools	• direct messaging system • following

Table 1. Main features of the SPLICE platform.

The initial prototype will require extensive testing during development and, subsequently, during the beta testing stage. Our objective is to involve as many scientists as possible in this process, and to integrate their feedback during the optimization of the platform. This process will occur repetitively in cycles in which new or improved features are released, feedback collected and the code optimized until no further improvements are possible from the technical and design point of view.

The release/feedback cycles will also be instrumental in collecting data about the various features of the SPLICE system, their practicality in real-life scenarios, and their acceptance by the scientific community. This data will then be used as the basis for reaching the second major objective – creation of a road-map towards web-native publishing and open science.

The advantages of the SPLICE concept over existing practices and potential impact

The proposed platform would integrate web-native publication with reviewing features and a reputation system. The system would not only solve most of the existing issues open science movement is trying to address, but also have much broader implications on science, science policy, funding agencies and economy.

The speed of communication would be dramatically improved. By breaking down the publication into smaller pieces, the publication lag would be practically eliminated. If data objects are communicated in near-real time, all results would be published and not only those that conform in a story.

With respect to measuring the contribution of individual scientists, publishing data objects would make it much easier to assign credit to individuals, because individual data objects will have fewer authors than paper-based publications have today. The author's contribution will be the basis of an elaborate but transparent reputation system, which would quantify precisely the contribution of each scientist, and break it down by activity. It will be possible to assess who does good experiments, who analyzes information well and combines data objects into sets and writes narrative articles, and who good reviewers are.

The latest point opens the room for thoughts about specialization of labor in the scientific system. Currently, the system is not efficient as most participants need to be skilled at all aspects of the scientific process. A PhD student is trained at doing experiments, analyzing data, writing papers, perhaps also writing grant proposals and reviewing, but only a small portion of trained scientists in the end reaches tenure and benefits from those skills. Recently, Alberts at al., using US as an example, argued that the current system in biomedical sciences is unsustainable and that the number of PhD students should be reduced, while the number of skilled and specialized staff scientists should be increased instead[15]. We agree with this view and believe that breaking up the scientific process into separate, but integrated parts, would provide the basis for this transition.

Further, opening of the review and integrating the reviewing activity into the reputation formula would open the possibility for scientists to continue contributing even after leaving academic research, as keeping their reputation high, and even increasing it, might be a sufficient incentive for them to remain active in their field of research as reviewers.

The overview of the SPLICE concept is given in Figure 3.

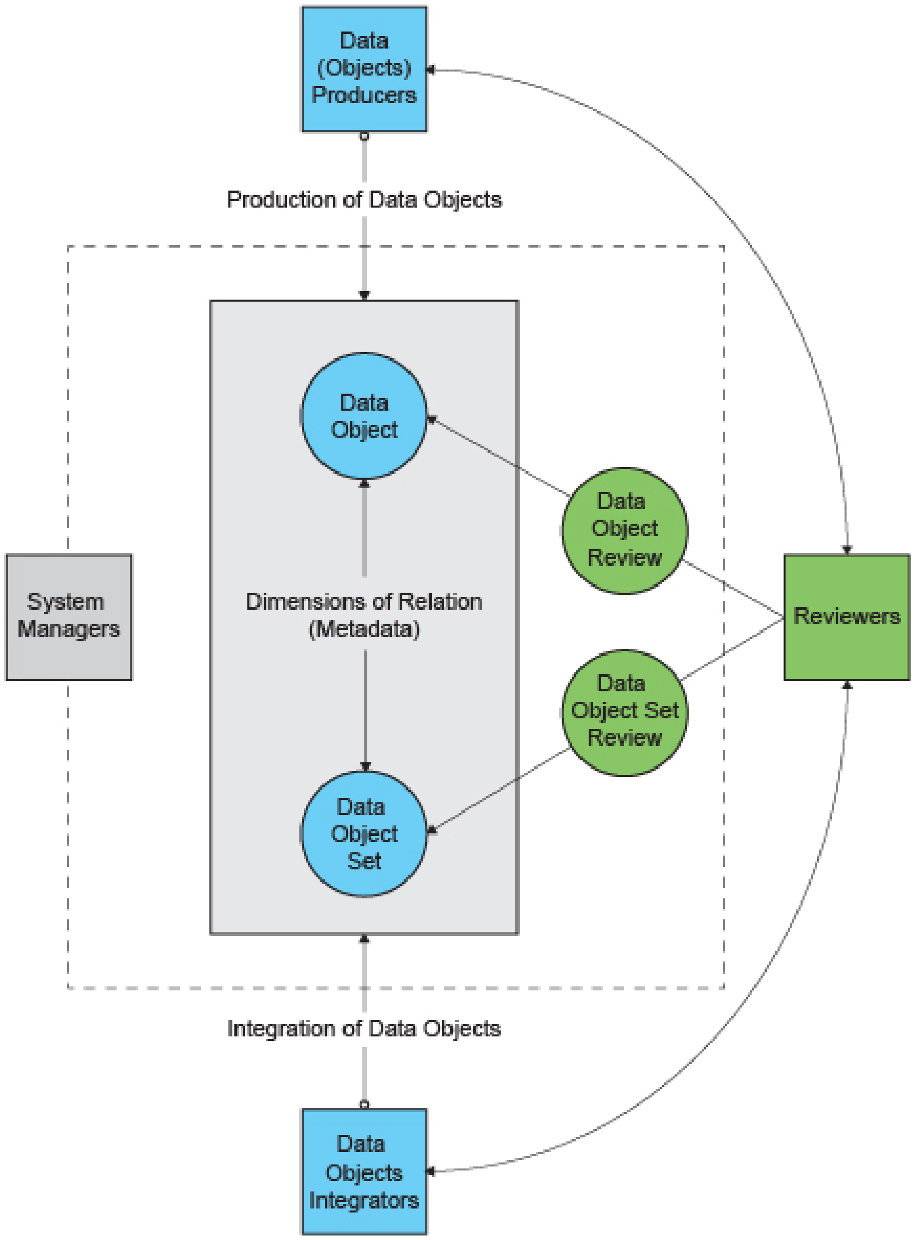

Figure 3. A scheme of the SPLICE concept.
Data objects are created by experimental scientists. They contain structured metadata and can be combined into sets of data objects (elaborated by narrative articles), by respecting the metadata values. Since metadata is used to determine whether or not two objects can be brought in relation to one another, we also call them dimensions of relation. Integration of data objects into sets can be done by the same scientists who created the data objects in the first place, but also by other colleagues, opening the possibility of data analyst/integrator positions which could become an important part of the research ecosystem. Both data objects and data object sets are subject to review. Reviewing is yet another line of work, which may be executed by experimental scientists, but also by specialized reviewers. Reviews are subject to community feedback as well to ensure their high quality. The entire system will require supervision and support of administrators.

It is conceivable that the proposed system could at later stages be integrated with work steps prior to creation of data objects (for instance, integration with electronic lab-books could make creation of data objects easier and to a certain degree automated), as well as work steps that follow publication (e.g. science communication to public).

From the technical point of view, the SPLICE concept carries some interesting advantages. For instance, in the current publication system it is very hard to eliminate errors from publications. Even if publications are retracted, numerous publications citing the original publication still remain in circulation. In an interconnected and highly structured system we propose, any type of change (either simple editing, adding a note or retraction) can be automatically propagated through the system to all objects that are in relation to the item in question.

Further, in addition to reviewing and scoring, interconnecting and reusing objects would provide a basis for much more accurate measurement of impact of research. It would suddenly be possible to calculate which data objects have been reused, and from the impact of the follow up work retroactively calculate the impact of the original research.

All the statistics would be calculated automatically at the level of content and scientists. Due to the structure of the database with respect to institutions and their hierarchy, which we intend to adopt from the Life Science Network project, it will also be possible to aggregate the statistics at the level of research groups, departments, institutes, or even entire cities and countries. This information could be very valuable for funding agencies like DFG.

This hierarchical infrastructure map would also help us deal with potential risks, such as attempts at gaming the reputation system. All scientists will be positioned uniquely within the system, and have a unique set of relations within it. By looking at the relations between scientists and their place of work, the content they publish etc., we will be able to distinguish reviews and ratings coming from scientists which overlap in some relations with the scientist they review, from those who don't. For example, if two scientists are authors of the same publication, or work in the same institution, when they mutually rate each other's work, those ratings might carry different weight than if a scientist with no co-authorships or working in a different institute does the same.

Since an idea can be considered a special form of a data object, the system we propose might favor sharing of ideas instead of withholding them. We believe that, if data objects consisting only of ideas and hypotheses are handled as other data objects (get a time stamp, a unique digital identifier and a handle), scientists would share their ideas and review ideas generated by other scientists. The implementation of this feature would allow for poor ideas to be filtered out in the first step of the research cycle and in that way reduce waste of resources on research which has conceptual flaws. Of course, the mindset of the community and all the stakeholders will have to change. This change might be stimulated if such data objects are also integrated into the reputation system and carry a significant weight when allocating research funding.

By accepting data objects as elementary publishing units and the idea that data object sets can be built from own, but also other researcher's objects, institutions and scientists which can't afford expensive equipment to pursue cutting edge experimental science could benefit enormously and contribute more significantly to the scientific knowledge base, because they could focus more on synthetic experimentation, which would be based on the sampling of experimental space consisting of data objects from researchers all over the world.

Moreover, it is easy to envision that an open and transparent web-based publication system, if pursued as suggested here, might significantly improve research efficiency and rationalize the use of research funds and, thus, have far reaching consequences for the whole research landscape as well. It is therefore important to approach the topic carefully and, following the development of a prototype, design a road-map towards implementation that will point to all the potential risks and needs for intervention.

O2: A road-map towards web-native publishing and open science

The implementation of open science objectives is confronted with many challenges. Those challenges are not of technical or financial nature, but predominantly social. It is therefore important to identify those issues as early in the transition process as possible and to integrate potential solutions with the development of technical tools. This is the second objective of this project. Our goal is to find out where intervention by policy makers will be needed to enable the transition to web-based publishing.

We plan to empirically asses the SPLICE concept by collecting extensive feedback and usage data from beta testers, statistically analyze it and present it in form of a road-map. The road-map will not only provide guidelines on implementation of technical tools and features, but also identify possible shortcomings in the concept and their potential impact on future scientific knowledge management and dissemination systems.

Transition to open science in some scientific disciplines has advanced more than in others. For instance, in physics, sharing of manuscripts through an open access repository (arXiv) for feedback prior to submission for publication in scientific journals is already a common practice[16]. When it comes to other disciplines, such as biomedical sciences, transition of this kind has had much less success, regardless of the availability of digital infrastructures. The reasons for this lie in the culture of the research community.

The system of values in biomedical research is focused predominately on publications in high-impact scientific journals. Some of those journals have become brands with reputation, and publishing in those journals is automatically seen as success and rewarded with jobs and grants. It is important to understand that in order for the SPLICE system proposed in this project to be implemented, we also need to adopt a new system of values. Our perceptions need to change as well, and this change will not occur on its own, but will need to be stimulated.

Even at the level of prototype building, the biggest risk will be to recruit beta testers that will be willing to invest time to study and use the system we create. However, we believe that at this initial stage, the community will ultimately respond positively. On a few occasions, some of the ideas presented here have been a topic of presentations in front of scientific audience and resulted in a lot of enthusiasm and interest.

Taking into consideration the resistance the stakeholders in the current publication system will provide, it is important to stress that the transition to open science will likely be a slow and sequential process. What seems inevitable, is that policy makers and funding agencies will play a major role in facilitating the transition towards open science by providing incentives and conditioning research funding in a way that stimulates change. Conclusions and the road-map that will result from this project will be instrumental in guiding policy makers and helping them manage this process.

2.3 Arbeitsprogramm und Umsetzung (Work programme and proposed research methods)

The project is organized around five work packages, which are summarized in Table 2. The first work package is focusing on project management and coordination, including organization of two workshops. Work packages 2 and 3 are carrying the bulk of activity and include development of the SPLICE prototype, as well as platform testing and optimization. The final work package deals with creation of the road-map towards implementation of open science.

Work package	Work package title	PM	Start month	End month
WP1	Project management activities	8	M01	M36
WP2	Development of the SPLICE prototype	48	M01	M18
WP3	Prototype testing and optimization	46	M19	M36
WP4	Road-map towards open science implementation	6	M30	M36
	TOTAL	108		

Table 2. Work packages of the SPLICE project.

Work package 1: Project management

The main objective of work package 1 is to ensure execution of all project tasks and delivery of promised results. This will be achieved through effective coordination, communication and feedback between all participants in the project. The project manager role will be assigned to Aleksander Benjak. Project manager will be responsible for direction of all project-related activities. Project manager will also monitor all legal, financial and administrative issues that might arise.

As part of this work package, we also plan to organize two workshops. The first workshop will be organized at the beginning of the funded period. The purpose of the workshop will be to: i) inform the community of the SPLICE project and thereby set the ground for the recruitment of beta testers (see WP3); ii) to collect initial ideas that will help us refine detailed specifications for the SPLICE platform (see WP2). The second workshop will be organized following the initial release of the platform prototype (beta phase). The purpose will be to explain the prototype to participants (individual researchers and representatives of institutions or consortia participating in beta testing) and collect first feedback on the developed prototype and its features.

Another task in this work package will be to recruit beta testers. The goal will be to promote the SPLICE concept and engage as many scientists as possible, who will help us in reaching the project's objectives by providing feedback. The project is going to be implemented in Heidelberg, which is a scientific hub and hosts a large university and several large research centers (e.g. German Cancer Research Center, European Molecular Biology Laboratory, Max Planck Institute for Medical Research). Those centers will be the starting point from which we plan to expand our recruiting activities to other German and international research institutions through visits, mailing campaigns and social media.

Task 1.1 (01-36): Administrative and financial management of the project
Task 1.2 (01-36): Overall planning and timing of project progress
Task 1.3 (01-36): Communication and day-to-day project coordination
Task 1.4 (01-03, 10-12): Organization of workshops
Task 1.5 (04-18): Recruitment of beta testers

Work package 2: Development of a prototype

The objective of this work package is to develop (program and design) the SPLICE platform. Given the complexity of the platform and underlying programming, the development will be split in several phases and some features will be based on the existing code provided by Life Science Network gGmbH. The goal is to establish a core version of the platform as early in the development process as possible and to add various functions in a successive and modular fashion. Upon completion of the most critical features and their successful integration, the platform will be launched (beta version).

The platform features are listed above in Table 1. However, as in a typical web-application development project, the development will start by creating detailed specifications sheets which developers can interpret and translate into code. These specifications will be refined with feedback of workshop participants.

Simultaneously with the creation of specifications, some ground programming work will start immediately. Obvious features and modules which can be adopted from the existing code provided by Life Science Network gGmbH will be implemented (registration, commenting module etc.). Also, wireframes and dummy templates, outlining the first look and providing a glimpse of user experience will be created by the end of the first six months into the project. Naturally, server-side infrastructure (hosting, data storage etc.) will be set up during this initial phase as well.

As the specifications are created, we will start programming the critical features: submission of data objects, search and integration of data objects into sets, version-tracking etc. Back-end development (database) as well as front-end development will run simultaneously. We estimate development of those features and front-end design to last 12 months.

In order to ensure the best possible implementation of this work package and work package 3, we will rely on web-technologies and infrastructure that allow agile development. The primary coding language of the platform will be Python. It will be coded using the Django framework[17] and hosted on third party servers (Heroku[18] and Amazon Web Services[19]) which will allow easy deployment of new code and provide tools for quick debugging, all while keeping the costs reasonably low.

Task 2.1 (01-06): Creation of detailed specifications for development
Task 2.2 (01-02): Set-up of server-side infrastructure
Task 2.3 (01-06): Adaptation of the existing code
Task 2.4 (01-06): Wireframes and mock-up pages
Task 2.5 (07-18): Implementation of specifications into a prototype
Task 2.6 (19-36): Beta release

Work package 3: Prototype testing and optimization by iteration

The platform will undergo two test phases. The goal of these tests will be to eliminate all bugs, system errors and evaluate and improve the functionality of the different features.

Initial assessment (alpha testing) of the features will be carried out internally. This step will be needed to evaluate the practicality of the platform and the different features, and eliminate all coding errors and obvious design mistakes. Alpha testing will occur simultaneously with the development of the platform.

Following the development of the initial prototype and the alpha testing phase, the platform will be opened to a larger group of scientists. Scientists will provide their feedback on the overall design and functionality of the different features. The feedback will be processed and the acceptance of different features will be assessed. During the beta testing phase, refinements of the platform and development of novel features will also take place in accordance with community feedback.

We estimate beta testing phase to start a year and a half into the project. Following experience with our previous projects, we believe it is best to launch a version of the prototype as early in the development process as possible, as user feedback can lead to significant changes. Therefore, improving the features in cycles typically lasts longer than the initial implementation.

Task 3.1 (07-18): Alpha testing
Task 3.2 (19-36): Beta testing and incremental optimization of the platform

Work package 4: A road-map towards implementation of open science

During the second half of the funded period, we plan to collect feedback on the platform and the developed features. The feedback will be analyzed, not only to optimize the developed prototype, but also to understand scientists' behavior and assess which features are likely to be adopted and which not. This will allow us to break down the workflow in different parts based on their acceptance and create a road-map towards a broad scale implementation of the SPLICE concept, which will include suggestions for new policies and incentives.

The feedback will be collected using several approaches. Firstly, workshops will be instrumental not only to discuss how to integrate certain features into the SPLICE platform, but also to collect first thoughts about potential issues. Secondly, once the platform is launched and opened to a broader group of scientists, we will gather information about their behavior using automated analytic tools (for instance Google Analytics). Such tools, as well as metrics which will be integrated directly into the platform, will help us identify the problems more precisely. Also, to understand why a particular problem occurs, we will launch periodic surveys. Results of those surveys will help us distinguish between the technical, design and user experience issues on one side, and cultural problems on the other. While one group of issues will be resolved through development iterations, the other will be addressed in the road-map, including suggestions on how the issues can be resolved.

Task 4.1 (19-36): Collection of site usage data and analysis
Task 4.2 (19-36): Surveys and analysis of the collected feedback
Task 4.3 (31-36): Creation of a road-map to implementation of open science

Timing of the Work Packages and their components - Gantt chart																																					
WPs	**Tasks**	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
WP1	T1.1 Administrative and financial management of the project	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
	T1.2 Overall planning and timing of project progress	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
	T1.3 Communication and day-to-day project coordination	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
	T1.4 Organization of w orkshops	■	■	■							■	■	■																								
	T1.5 Recruitment of beta testers				■	■	■	■	■	■	■	■	■	■	■	■	■	■	■																		
WP2	T2.1 Creation of detailed specifications for development	■	■	■	■	■	■																														
	T2.2 Set-up of server-side infrastructure	■	■																																		
	T2.3 Adaptation of the existing code	■	■	■	■	■	■																														
	T2.4 Wireframes and mock-up pages	■	■	■	■	■	■																														
	T2.5 Implementation of specifications into a prototype							■	■	■	■	■	■	■	■	■	■	■	■																		
	T2.6 Beta release																			■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
WP3	T3.1 Alpha testing							■	■	■	■	■	■	■	■	■	■	■	■																		
	T3.2 Beta testing and incremental optimization of the platform																			■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
WP4	T4.1 Collection of site usage data and analysis																			■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
	T4.2 Surveys and analysis of the collected feedback																			■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
	T4.3 Creation of a road-map to implementation of open science																															■	■	■	■	■	■

Table 3. Gantt chart for the SPLICE project.

2.4 Maßnahmen zur Erfüllung der Förderbedingungen und Umgang mit den Projektergebnissen

The prototype of the platform will be developed with the goal to research and establish a workflow for publishing in a web-native format and optimize it according to community feedback. Community reaction and acquired knowledge will be used to create a road-map for implementation of open science. The conclusions will be presented and disseminated in the scientific community through presentations and articles in scientific journals and online platforms.

In addition to creating a road-map, we will make sure the platform can continue to operate beyond the funding period. The task of platform operation and maintenance will be taken by the Life Science Network gGmbH, our partner on the project.

Based on the experience of Life Science Network gGmbH which currently maintains two internet platforms (www.lifescience.net and www.scimplified.com), we will apply the same kind of strategy to the SPLICE project. During development, we will take into consideration and comply with existing standards wherever possible. The programming language and the technologies used to develop the platform will be standard to ensure easy adaptation and scalability. This will make it possible to launch a production-level platform following the completion of the project, should the feedback of the community be positive.

As part of the platform, we will also develop a set of web application programming interfaces (APIs). Those will allow external access to the resources of the SPLICE platform and facilitate their integration in other platforms or software tailored to retrieve various data and statistics, or take advantage of the SPLICE web-native publication format. The source code developed for the purposes of the project will be made open and shared with the public through a standard code-sharing repository (e.g. GitHub).

2.5 Erläuterungen zur inhaltlichen und finanziellen Projektbeteiligung von Kooperationspartnerinnen und Kooperationspartnern im Ausland

No foreign cooperation partners will contribute to the project in terms of development or financial contribution. However, we plan to involve partners as beta testers of the prototype platform we aim to develop.

3 Literaturverzeichnis (Bibliography)

1. San Francisco Declaration on Research Assessment. at <http://www.ascb.org/dora/>

2. Yarkoni, T. Designing next-generation platforms for evaluating scientific output: what scientists can learn from the social web. *Front. Comput. Neurosci.* **6,** 72 (2012).

3. Glenn Begley, C. & Ellis, L. M. Raise standards for preclinical cancer research. *Nature* **483,** 531–533 (2011).

4. Figshare. at <http://figshare.com/>

5. Altmetrics. at <http://www.altmetric.com/>

6. Impactstory. at <https://impactstory.org/>

7. Rubriq. at <http://www.rubriq.com/>

8. F1000. at <http://f1000.com/>

9. Bohannon, J. Who's Afraid of Peer Review? **342,** 60–65 (2013).

10. Rooyen, S. Van, Godlee, F., Evans, S., Black, N. & Smith, R. Effect of open peer review on quality of reviews and on reviewers ' recommendations : a randomised trial. **318,** 23–27 (1999).

11. Kriegeskorte, N., Walther, A. & Deca, D. An emerging consensus for open evaluation: 18 visions for the future of scientific publishing. *Front. Comput. Neurosci.* **6,** 94 (2012).

12. Mons, B. *et al.* The value of data. *Nat. Genet.* **43,** 281–3 (2011).

13. Scientific data. at <http://www.nature.com/sdata/>

14. Minimum Information for Biological and Biomedical Investigations (MIBBI) Project. at <http://mibbi.sourceforge.net/legacy.shtml>

15. Alberts, B., Kirschner, M. W., Tilghman, S. & Varmus, H. Rescuing US biomedical research from its systemic flaws. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 5773–7 (2014).

16. ArXiv. at <http://arxiv.org/>

17. Django. at <https://www.djangoproject.com/>

18. Heroku. at <https://www.heroku.com/>

19. Amazon Web Services. at <http://aws.amazon.com/>

4 Finanzierung des Vorhabens (Project funding)

4.1 Beantragte Module/Mittel (Requested modules/funds)

Funding for staff

We apply for the two positions (postdoctoral level) for Aleksander Benjak (Felix Wieland). Both positions should run over the entire duration of the proposed project.
(63.600 EUR x 2 staff x 3 year = 381.600 EUR)

One staff member will be in charge of project coordination, platform testing, communication with scientists, collecting of feedback, creation of specifications for development and other tasks.

The other position will be used to hire a front-end developer. Tasks will include design and development of the user interface for the project. The two staff will work together and collaborate closely with the third staff member, who will be hired by Life Science Network gGmbH and be responsible for back-end development and programming (see below).

Funding for travel

We apply for 3000 EUR for traveling expenses for each year. The money will be used to cover expenses of short traveling trips within Germany to give seminars at German research institutions and universities and participate in events where the project can be promoted and beta testers recruited.
(3000 EUR x 3 years = 9000 EUR)

Funding for work-shop organization

We apply for 24.000 EUR to organize two workshops. The events will be organized for 50 people.
(12.000 EUR x 2 events = 24.000 EUR)

Other expenses

For our partner, Life Science Network gGmbH, we request 80.000 Eur per year over the entire duration of the proposed project. The funds will be used to hire a back-end developer/programmer whose tasks will include setting up and maintenance of the server-side infrastructure, back-end development (design and development of the database) and related programming activities.

The funds will also cover platform hosting and data storage expenses for the period of three years during platform development.

	Personnel	Travel	Workshops	Hosting	**Total**
AB (Uni HD)	381.600	9.000	24.000	0	414.600
LSN gGmbH	228.000	0	0	12.000	240.000
Total	609.600	9.000	24.000	12.000	**654.600**

Table 4. Funding overview for the SPLICE project.

4.2 Eigenleistung (Own contribution)

The group of Felix Wieland will host the project and provide the necessary office space and computer equipment for two positions.

Life Science Network gGmbH will provide the office and computer equipment for the third position. Further, Life Science Network gGmbH will give access to the source code of the Life Science Network and Science Simplified projects, in an effort to speed up development of the SPLICE platform.

5 Voraussetzungen für die Durchführung des Vorhabens (Project requirements)

5.1 Angaben zur Dienststellung (Employment status information)

Aleksander Benjak, Dr. rer. nat.
Academic co-worker (Akademischer Mitarbeiter) in Wieland Group (until December 2014)
Biochemistry Center of the Heidelberg University, Heidelberg

The position is funded by Federation of European Biochemical Societies (FEBS).

5.2 Zusammensetzung der Projektarbeitsgruppe (Composition of the project group)

Aleksander Benjak (Felix Wieland, BZH Heidelberg)
The group of Felix Wieland provides an excellent environment for the execution of the project. While engaging in experimental research, Felix Wieland is also the Managing Editor of FEBS Letters. In this capacity he has gained valuable insights into current publishing trends and is knowledgeable of the whole publication process, including scientific evaluation, editing and post-acceptance processing of manuscripts. This will expose the project both to the scientists and the publishing world and ensure continuous feedback.

Life Science Network gGmbH
The non-profit organization Life Science Network gGmbH has been devoted to developing high-end web applications for scientists since 2011. SPLICE will be the third project LSN gGmbH will engage in. The project will benefit from the experience gained through the first two projects, as well as directly profit from the provided code.

Life Science Network gGmbH is located in Heidelberg, in close proximity to BZH Heidelberg. This will enable effective communication of all the team members and regular meetings.

The organization is represented by Dr. Alen Piljic, who contributed significantly to the creation of the SPLICE concept, together with Dr. John Lock (Karolinska Institute).

5.3 Zusammenarbeit mit anderen Institutionen und anderen Wissenschaftlerinnen und Wissenschaftlern

5.3.1 Institutionen oder Wissenschaftlerinnen und Wissenschaftler, mit denen für dieses Vorhaben eine konkrete Vereinbarung besteht

The SPLICE project will be executed in cooperation with Life Science Network gGmbH, as described in section 5.2.

5.3.2 Institutionen, Wissenschaftlerinnen und Wissenschaftler, mit denen in den letzten drei Jahren gemeinsame Projekte durchgeführt wurden

None.

5.4 Erklärungen zur Erfüllung der Förderbedingungen

Hereby we declare the following:

- All results and publications resulting from the project and any relevant documentation will be available via open access, making them widely accessible for use by third parties.
- Source code for the software developed under the project will be documented in accordance with the principles of open source and made available for use by third parties.

5.5 Projektrelevante Zusammenarbeit mit erwerbswirtschaftlichen Unternehmen

None.

5.6 Projektrelevante Beteiligungen an erwerbswirtschaftlichen Unternehmen

None.

6 Ergänzende Erklärungen (Additional information)

No other funding proposals for this project have previously been submitted to a third party.