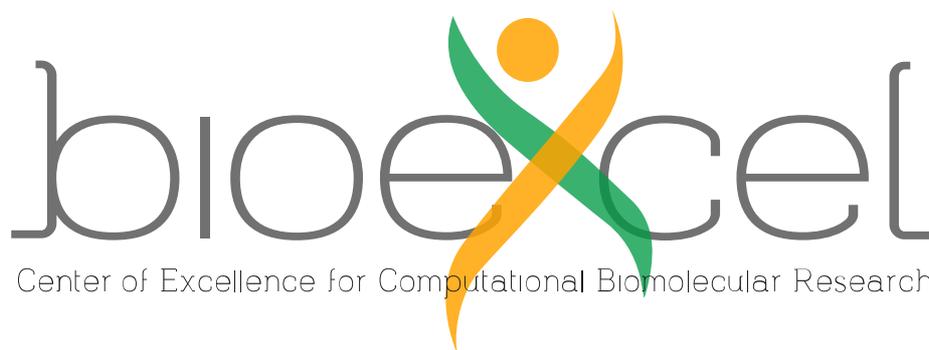


H2020 EINFRA-5-2015



www.bioexcel.eu

Project Number 675728

D5.3 – Updated business plan

WP5: Governance and Sustainability



Copyright© 2015-2018 The partners of the BioExcel Consortium

Document Information

Deliverable Number	D5.3
Deliverable Name	Updated business plan
Due Date	2018-04-30 (PM30)
Deliverable Lead	FT
Authors	Holger Noack (FT), Ian Harrow (IHC), Alexandre Bonvin (UU), Adam Carter (EPCC), Rossen Apostolov (KTH), Erwin Laure (KTH)
Keywords	Sustainability, users, support, services, commercial
WP	WP5
Nature	Report
Dissemination Level	Public
Final Version Date	
Reviewed by	PMB
MGT Board Approval	2018-04-30

Executive Summary

This is the third iteration of our business plan. It reflects the progress of the sustainability planning and the commercial intentions of the consortium as of April 2018.

Section one gives a brief presentation of our mission and our sustainability goals, followed by a description of the biomolecular researchers that we regard as our main users and potential clients.

In section two, we estimate the size of our current academic user community to be at least 15.000 active users. We provide findings from surveys among academic and industrial users, showing that usability aspects and compatibility of software are major concerns. We further present data from HPC market studies, based on which we estimate that the European market for biomolecular HPC software has a size of ca €170 million in 2018.

In section three we present the business model canvas for the BioExcel CoE, that will be discussed in greater detail throughout the following sections.

Section four shows our value proposition that consists of support for the development of software & workflows, provision of demonstrators, as well as support aimed at increasing the competency of researchers using biomolecular software.

We discuss our commercial offerings in section five, where we have the intention to develop online training and workflow-related services/products. The product definition is still in progress, with prototyping, provision of demonstrators and product testing in focus groups to be done during the remaining months. The goal is to have the first prototyped offerings in place by the end of this project.

The marketing strategy is discussed in section six, mentioning online and offline channels that we are using, such as e.g. Twitter or networking events. We also present our sales efforts that include usage of a CRM tool to follow the progress on our leads.

Aspects of intellectual property are discussed in section 7, foremost touching on the license situation of our codes. The conclusion is that we can use the BioExcel-supported software in commercial settings.

In section eight we present the envisioned BioExcel structure, which will consist of a consortium and a social enterprise. The social enterprise will be our single interface for commercial activities, while the consortium will work on academic aspects that are financed through public grants.

The financial management is touched-on in section nine, showing that we will have three types of revenue to fund software development and provision of user support.

Where appropriate, sections conclude with action points, indicating next steps and the corresponding timelines.

Contents

1	<u>OVERVIEW</u>	5
1.1	VISION AND MISSION	5
1.2	SUSTAINABILITY SUCCESS GOALS	5
1.3	DESCRIPTION OF USERS & POTENTIAL CLIENTS	6
2	<u>BIOMOLECULAR RESEARCHERS</u>	7
2.1	SIZE OF THE ACADEMIC RESEARCH COMMUNITY	7
2.2	SURVEY FINDINGS FOR ACADEMIA	8
2.3	SIZE OF THE INDUSTRIAL RESEARCH COMMUNITY	9
2.4	SURVEY FINDINGS INDUSTRY	9
2.5	MARKET SIZE FOR COMPUTATIONAL SOFTWARE	10
3	<u>BUSINESS OPPORTUNITIES THE BIOEXCEL COE</u>	10
3.1	BUSINESS MODEL CANVAS	11
4	<u>VALUE PROPOSITION</u>	11
5	<u>COMMERCIAL OFFERINGS</u>	13
5.1	CRITERIA AND SELECTION PROCESS	13
5.2	TIMELINE FOR PRODUCTIZATION	14
5.3	TRAINING PRODUCTS	14
5.4	WORKFLOW PRODUCTS	15
5.5	OTHER OFFERINGS	16
6	<u>MARKETING STRATEGY</u>	16
6.1	BRANDING	17
6.2	ONLINE MARKETING	17
6.3	OFFLINE MARKETING	18
6.4	TIMELINE	18
6.5	SALES STRATEGY & LEAD GENERATION	19
7	<u>IP MANAGEMENT</u>	19
7.1	DESCRIPTION OF SOFTWARE LICENSES	19
7.2	COMMERCIAL ASPECTS	21
8	<u>LEGAL STRUCTURE</u>	21
8.1	OPERATIONS MANAGEMENT	22
8.2	MODUS OPERANDI	22
9	<u>FINANCIAL MANAGEMENT</u>	23
9.1	SOCIAL ENTERPRISE STARTUP FUNDS	23
9.2	SOCIAL ENTERPRISE OPERATIONAL FUNDS	24
9.3	SOCIAL ENTERPRISE REVENUE MANAGEMENT	24
	<u>APPENDIX 1: TIMELINE AND ACTION POINTS</u>	24

1 Overview

1.1 Vision and mission

Computational biology/chemistry, as a research discipline, has become indispensable to address societal challenges in health, food supply and environment. Successful application of this research discipline is directly linked to its ability of generating reliable/accurate and meaningful results. However, computational methods have increased in complexity and variety during the last few decades, while trends in hardware development provide new possibilities to address computationally demanding research questions. As a consequence, for the individual researcher, it has become increasingly difficult to choose the most appropriate route for tackling a given research question, i.e. the choice of software, method, parameters and hardware environment to be utilized.

Within this context, BioExcel envisions becoming a facilitator for the exchange of expertise among computational researchers and software developers. The declared intention is to increase the quality of the biomolecular computational research that is performed by thousands of researchers worldwide, and in doing so we will also accelerate the industrial application of the research in this field. To achieve this, BioExcel will (i) actively work to improve frequently used computational tools, and (ii) educate researchers in the best use of these computational tools.

Vision: *“A central hub for biomolecular modelling and simulations”*

Mission: *“Enabling better science by improving the most popular biomolecular software and spreading best practices and expertise among the communities.”*

1.2 Sustainability success goals

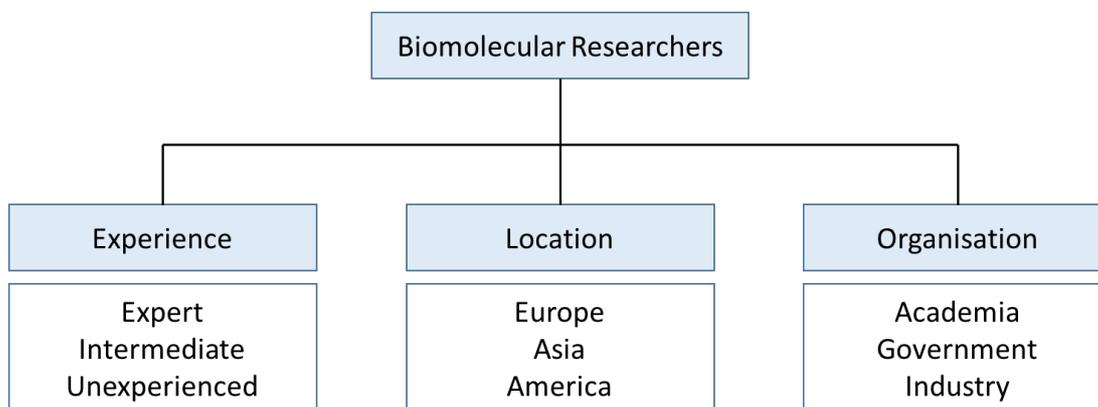
We have defined success goals that will contribute towards BioExcel’s mission as summarized in the table below. These success goals have been discussed with all partners to ensure broad support and alignment of the center’s activities.

Sustainability Goal	Short description
User community	Our user community should ideally consist of individuals with varying levels of experience, and from different types of research environments. Community members should feel encouraged to actively contribute towards BioExcel’s mission.
Expertise sharing	Facilitate sharing of expertise among different stakeholders, support competence matching, encourage collaboration.
User training	Provision of training that enables users to make best use of the available software and the possibilities of existing hardware.
Consultation	Provision of case- and user-specific support, either through collaborative research projects, or as commercial offering.

State-of-the-art software	Constantly improve ease-of-use, scalability, and efficiency of important software, taking into account user feedback.
Workflows	Provide workflow components and entire workflows to e.g. increase interoperability of software tools, user-friendliness or efficiency.
Branding	Establish professional online presence and actively work with marketing to improve visibility, attract community members, competence, and potential customers.
Financials	Establish diverse revenue streams to become less dependent on recurring public funding.

1.3 Description of Users & Potential Clients

The emphasis of the center’s activities is placed on the end-user, i.e. the biomolecular researchers that use software and hardware to perform calculations with the overall goal to solve biomolecular research challenges. These research users can be further segmented by their level of experience, location, and type of host organisation.



The level of experience is especially important when designing workshops or training programs. BioExcel has developed proto personae that reflect different granularities of user experience, and has performed a gap analysis, to determine the training needs of these different personae. We are currently in the process of refining these to be full user personae. All of this is described in detail in the deliverables from the training work package.

Differentiating researchers by type of host organisation or research environment provides some information about specific needs. In general, academia has a broad support of qualified post-graduate students and postdocs, who usually are accustomed to open-source software. Academic researchers are thus more likely to test new functionalities and explore the boundaries of existing software solutions. But academia also has educational responsibilities, which will affect support needs. This may include practical aspects in delivery of training, such as a recurring need to access remote computing infrastructures in a timely manner and deploy virtual machines for a number of largely inexperienced users. In comparison, researchers at governmental organisations do not have educational obligations, and tend to work more towards applied research. These researchers are thus more likely to require stable tools that excel in performing repetitive tasks. Finally, researchers from industry are often required to observe regulatory constraints, such as e.g. the use of software with version control to

validate production of data packages. In interviews, these researchers reported on extreme time pressure to achieve results, which leaves them little time to “experiment” with open-source software that often requires studying extensive manuals. Instead, within industry, there is preference for validated and well maintained “golden-standard” solutions, as well as an acceptance to use consultancy services for customized support and without risking disclosure of intellectual property (IP).

2 Biomolecular researchers

BioExcel’s community consists primarily of users of the software adopted and further developed in BioExcel, but also of researchers interested in our newly developed workflows and virtual machines. In the first instance, we will focus on users of the codes that form the largest section of our user base. Regular surveys among users of GROMACS and HADDOCK show that these users, in their overwhelming majority, work in non-profit/academic environments located in various regions worldwide.

	GROMACS¹ 224 responses	HADDOCK² 755 responses
User type	94% non-profit 6% for-profit	97% non-profit 3% for-profit
Location (3 largest)	50% Europe 19% North America 16% Asia	23% India 21% Europe 19% North America

2.1 Size of the academic research community

The size of BioExcel's academic research community can be estimated by the number of registrations, which is possible in the case of HADDOCK, or from the number of publications that refer to the biomolecular software used.

Over the past five years, the number of registered users on the HADDOCK web portal has constantly increased. In April 2018, the total number of registered users on the HADDOCK web portal exceeded 10.800³. In addition, there are circa 1000 local installations of the HADDOCK software, plus some installations accessible to the 103 member organisations of the SBGrid consortium. Since not all of the registered users are at all times active, it is good to also look at the number of new registrations and active users.

HADDOCK web portal	2015	2016	increase
# of registered users	6699	8320	24%
# of new registrations	1450	1621	12%
# of unique users submitting at least 1 run	1262	1315	4%
# of dockings runs	24806	28402	14%

Table: HADDOCK web portal use for the years 2015 and 2016, showing the total number of registered users, and for comparison the number of new registrations and active users.

¹ GROMACS feature and usage survey, 2014

² HADDOCK online survey, access on April 2017

³ <http://haddock.science.uu.nl/services/HADDOCK2.2/haddock.php>

The table below shows the number of articles that either cite or mention one of the BioExcel codes. In this table PMX is missing, since the software has only very recently been made available for research community.

YEAR	CPMD	GROMACS	HADDOCK
2000	89	20	-
2001	103	33	-
2002	150	111	-
2003	168	177	2
2004	209	332	35
2005	284	581	63
2006	316	725	77
2007	345	952	96
2008	412	1198	119
2009	416	1262	166
2010	397	1478	211
2011	449	1863	248
2012	473	2149	278
2013	440	2672	356
2014	479	3275	340
2015	384	3335	330
2016	336	4038	385

In 2016, we had 1315 active users submitting at least one calculation through the HADDOCK web portal. This compares to 385 publications referring to HADDOCK during the same year. Hence, we can assume that in the case of HADDOCK, each publication represents on average 3 active users. It is not unreasonable to apply this usage/publication ratio to the other BioExcel software. Based on this approach, we can estimate that the academic user community consisted of at least 15.000 active researchers in 2016.

Action point	Due
Update the table of references and the number of HADDOCK web portal users with fresh numbers from 2017.	June 2018

2.2 Survey findings for academia

In the case of HADDOCK, more than 40% of the survey respondents use the software for less than a year, which is not surprising given the growth of new registrations. This means, however, that there is a big need for training and support, which currently is delivered through tutorials, and online support. Users are in general very satisfied with the quality of the workshops, and when asked, more than 70% would accept paying a small fee if a workshop is hosted in a nearby institute. The situation is similar in the case of GROMACS. Users appreciate the performance of the software and the support that is provided, but they would like to see even more extensive tutorials and more occasions to attend workshops.

Regarding the software, most users made it clear that apart from its performance, the free access of the software is of crucial importance. Besides further improvements in performance and efficiency, the three most common requests are:

1. Integration of analysis tools
2. Compatibility of software with more data types and other software
3. Provision of tools that support building molecular systems and setting up calculations

2.3 Size of the industrial research community

From the surveys, we estimate that between 3% to 6% of our current user community works at for-profit organizations. We do not know what they use the software for, or how frequently they are using it. There is certainly potential to grow our reach out to industrial researchers and thereby promote the software that is supported through BioExcel.

To set a target based on the accessible number of industrial users is rather difficult. One could count the number of large pharma companies, estimate the average number of employees that work with computational biomolecular research, and extrapolate. But according to numbers from the OECD, 60-70 per cent of jobs are located in small enterprises. Furthermore, we would miss-out on non-pharma companies that have a biomolecular component, such as e.g. agriculture or environment.

Just to get a first idea, we looked at job postings for “computational biologist” at large global job sites. We find that on average 20% of these postings are located within industry. Applying this ratio to the BioExcel community, would correspond to circa 3000 individuals.

Action point	Due
Refine the estimate of the number of industrial researchers by accessing publicly available data on industry sectors (e.g. CORDIS)	June 2018

2.4 Survey findings industry

Although industrial researchers represent only a small fraction of current users of BioExcel software, they are important being the most likely source of paying customers. In order to gain insights into their needs, we conducted direct interviews and site visits. The industrial researchers that we spoke with work with virtual screening, antibody engineering & lead optimization. They prefer to use industry-standard software from e.g. Schrödinger (Maestro) or the Chemical Computing Group (MOE). They further stated a preference to use internal computing resources, mainly for data security reasons. The challenges that they see are less related to scalability, but more about the quality of prediction and usability aspects:

1. Conversion between formats
2. Integration between software tools
3. Building and setting up systems, in particular:
 - a) Embedding molecules in membranes

- b) Finding optimal parameters for force fields
- c) Adding ligands, which often affects a) and b) above

Industrial researchers are aware of open-source software like GROMACS, but the complexity of the software can deter them, since it would require extensive read-up of literature to be able to use it properly. In general, there is a high pressure to shorten time-to-results, which leaves industrial researchers little time to evaluate different solutions that are deviating from a deploy and operate modus.

In principle, there is an interest to participate in use cases with BioExcel, on topics related to e.g. free-energy calculations or membrane chemistry. Some industrial researchers further signalled interest in potential service offerings such as web-training, workshops or site-visits from BioExcel computational experts.

These points have been factored into the development of our commercial offerings. Prototyping is underway for site visits, webinars, and workshops.

2.5 Market size for computational software

Simulations of biomolecular systems require effective computing techniques to achieve timely results. In their majority, computations are either performed as tightly coupled parallel jobs on high-performance computing infrastructures (HPC), or as loosely-coupled tasks on distributed networks of processors known as high-throughput computing (HTC).

The market size for computational biomolecular software solutions can be estimated by looking at HPC market studies that the International Data Corporation (IDC) has performed on behalf of the EC. According to the IDC market report from 2014-2015, HPC software generated global revenues of €4.4 billion in 2013, of which Europe's spending was estimated to €1.2 billion. It was further estimated that the European spending in this market will expand to €1.7 billion by 2018. Biosciences are the second largest HPC application segment standing for circa 10% of all spending in this sector. Hence, we can conclude that that the 2018 European market size for biomolecular computational software is around €170 million.

The market for software for molecular simulations and modelling is highly competitive with a range of offerings from many academic groups and commercial providers. As discussed in section five below, we currently intent to base one of our first commercial offerings on workflows. In this regard, we need to extend the market analysis to better understand the competitive landscape and opportunities, and how we can differentiate ourselves.

Action point	Due
Extend market analysis to understand opportunities for workflow-related offerings	October 2018

3 Business opportunities the BioExcel CoE

Quite a few mature commercial providers e.g. Schrödinger, Chemical Computing Group and Biovia 3DS offer a suite of products and services largely to a customer base

of industrial researchers. In contrast, non-profit researchers in their majority make use of freely available open-source software. The commercial solutions have the advantage that they are easy to use and come with support. Open-source software has the advantage that it can be customized to fit anyone's needs, given that a user has the expertise how to do that. And while not having a service-level of support, there is a broad community support.

The BioExcel CoE has the advantage that it is combining the somewhat fragmented communities of three important open-source software solutions under one roof. This accumulation allows to align efforts in software development and research, which is attractive for academic and industrial researchers alike. The whole economic boom related to new pharmaceuticals, such as e.g. immunotherapies for cancer treatment, is fuelling this opportunity area across the globe. We are seeing interest in BioExcel from as far as India, Singapore, USA and Australia, and therefore the centre needs to pin down its business model in order to be able to exploit its leading position and ensure sustainability for the future.

3.1 Business Model Canvas

Key Partnerships	Key Activities	Value & Services	Customer Relationships	Customer Segments
European Commission	Software development	De-facto quality brand	Research focuses	Academia
Universities	Provision of training and expertise	Training & Webinars	Guidance on code usage	Governmental entities
Research Communities	Key Resources Software and workflow developers Computational experts	Development of new workflows and tools	Channels Website and online platforms Social Media Events	Pharma
Industry		Steering development of important open-source codes		SMEs working with Biotech
Cost Structure		Revenue Structure		
Research staff Travelling Administration		Use of computing infrastructure Promotion & marketing Research grants Income from training events Income from consultancy, e.g. customisation or SimulationaaS		

4 Value proposition

BioExcel's value proposition is designed to meet the above-described user needs. We support software development of biomolecular codes to improve efficiency, ensure stable performance on modern hardware and enhance usability. In parallel, we also work to enhance user experience and productivity by developing efficient and extensible workflow solutions for automation of large scale computational tasks, which are becoming more and more the norm when using large scale (pre-Exascale) resources. In addition, taking full advantage of the provided tools requires considerable skill set, which BioExcel provides in the form of user support, training events and consultancy activities.

As part of our sustainability planning, some offerings are going to be provided in a commercial form, and those are described in the next section. The table below summarizes the current BioExcel efforts that are providing value to all users.

SOFTWARE	Description
HADDOCK GROMACS CPMD PMX	Improving efficiency, performance and usability. Software developers take user feedback into account during the development process. Through BioExcel, user needs are actively and continuously gathered through several use cases and interest groups. A detailed description of the software development can be found in WP1 deliverables.
Workflow components	Quality workflow components and wrappers for virtual machines are being developed to improve interoperability of different software tools, to simplify conversion between formats, and provide possibilities of designing efficient workflows for common or difficult research tasks. A detailed description of this development is provided in WP2. Workflows are also prototyped in use cases, which are described in WP3.
Demonstrators	BioExcel is demonstrating the quality of predictions that are possible with the BioExcel-promoted software through provision of benchmarks and through participation in challenges.

The table below gives an overview of the free support that is provided with the aim of increasing the competency of biomolecular researchers in using the codes.

COMPETENCY	Description
Website	Information hot-spot for biomolecular researchers to find tutorials for BioExcel software, get information about conferences, training offerings etc.
Training	Tutorials are provided in the form of use cases. A unified presentation shall make it easier for biomolecular researchers to move between different codes.
Knowledge resource center	Users can find training resources that are relevant for their level of experience. Example personas are being developed that will help researchers to identify exactly what competencies they need to acquire in order to progress their computational abilities.
Webinars	Speakers from BioExcel and from partner organisations are giving talks on the latest research topics, or present useful software functionalities. These events allow for Q&A sessions.
Workshops	Workshops for BioExcel-supported software are provided several times per year.
Support forum	Developers and computational experts support research users on specific questions regarding usage of BioExcel software.

5 Commercial Offerings

BioExcel aims to provide commercial offerings in order to become less dependent on public funding for the provision of user support and efforts in software development.

5.1 Criteria and selection process

The BioExcel partners met to discuss possible commercial offerings that can be provided to meet the above-described user needs. The goal was to identify services or products that are feasible, i.e. that we are able to provide based on the competencies that are available among the partners while taking limitations into account (dependencies, resources needed). The table below summarizes the possibilities that were considered.

	Comments	Feasibility	Timeframe
Software Spin-out			
HADDOCK, GROMACS	Dedicated development effort needed	High	Long
PMX, CPMD	Dedicated development effort needed	High	Long
Workflows & Libraries	Available	High	Short
Support			
Community as a service	Needs strong brand	Low	Long
24/7 support	Resource intensive	Low	Short
Consultancy	Performed at client site, resource intensive	Medium	Long
Contract research	Performed at BioExcel site, resources are more flexible	Medium	Long
Platform as a service			
HPC access	Needs secure environment	High	Long
Large-scale analysis	Needs secure environment	High	Long
Training			
Online	Needs access to infrastructure	High	Short
Webinars		High	Short
Face-to-face	Resource intensive	Low	Short

Among the proposed services, we tried to identify those that can be prototyped within 2018, and that do not require any EU funding for implementation. Ideally, these offerings should be scalable, and should have a short time to market.

Development of commercial software/add-ons would require a dedicated development effort, and has a long timeframe. In addition, it would require taking in investor capital to finance the software development, since it is unlikely that we can use the currently provided public funding for that. Consultancy is an interesting service especially for industrial users. But at the moment with our current BioExcel structure it has a long

lead time, since we do not have BioExcel experts available that could take-on eventually incoming assignments on-short notice. BioExcel could also act as a platform to enable partners that have the capacity to sell HPC access. This service would have a low workload requirement, but it would not align with the user needs as much as other service from the list.

Based on these considerations, we decided to focus our initial commercialisation efforts on training and workflows. The existing training offerings are well received among our current users. And the newly developed workflow-prototypes also had a positive reception in our use cases. In addition, industry representatives indicated that workflows could be interesting also from their perspective. Several BioExcel partners came forward as champions for each of the commercial offerings. EMBL-EBI will be leading the pilot training services whilst IRB will be leading workflows for the sustainability pilot.

5.2 Timeline for productization

The partners decided on a timeline until end of the project to take the two chosen products to a pre-competitive mature state.

2018	Training	Workflows
February March	Inventory of training modules and breakdown of effort needed to provide a commercial offering.	Inventory of workflows to identify the ones with commercial potential. Take input from use cases, IGs, and site visits into account.
April	Send out surveys to industry and academic users.	Gather Industry input on e.g. quality of libraries, choice of demonstrators
May June	Create content Create brochures and price-list Initiate promotion	
July August September	Test Integrate input from first test Test refined version	
Oct.	Collect feedback & finalize business plan	
(Nov.)	Community forum / product launch	

5.3 Training products

The inventory of training modules is concluded, and the top candidates are modules related to HADDOCK, and “Intro to HPC for Life Scientists”. We decided to start with HADDOCK, since we can start from the existing workshops that are appreciated by many researchers, but that due to limited resources are delivered only a few times a year. Our intention is to generate an online course that can be delivered repeatedly, to a large audience. Paying user could be offered interactive support, and in case of successful completion an official certificate. At the current stage, we intent to build on

material developed for the BioExcel Summer School in June 2018. In addition, we need to evaluate the possibilities to provide access to the HADDOCK web-portal for participants, which should be possible, and which would be another advantage of choosing HADDOCK. If this first commercial online course is successful, we will successively expand the catalogue to include courses for HPC, GROMACS and other BioExcel-supported software.

Pricing of courses is difficult. We expect that the majority of users for the HADDOCK course will stem from academia. Prices thus need to be feasible. Online courses have the advantage that they can be delivered at a lower fee than physical courses, and most importantly, they enable many users from any location to participate (only 21% of the HADDOCK users are located in Europe). To arrive at an attractive proposition, we intent to send out a survey to the BioExcel community in May 2018, to gather preferences on contents and format, and to test the acceptance for fees on different packages propositions.

In addition to the discussed offering of an online course, we have just now in April 2018 provided a full day of training to an industrial user that purchased a HADDOCK license. After this training, the company indicated interest for consultancy and for possible collaborative projects. It is our intention to investigate how training could be bundled with software licenses to be especially appealing to industrial users of the BioExcel supported codes.

Action point(s)	Due
Detailed definition of training product business model and competitive analysis to define a pricing strategy.	October 2018
Demonstrate what income can be generated from each activity, and discuss income targets.	November 2018

5.4 Workflow Products

The inventory of the workflows has resulted in the selection of a workflow for protein mutations (pymdsetup). This workflow was made public at GitHub in order for researchers to test it and as a pre-marketing action to attract interest.

The aim of the consortium is to sell consultancy services based on customisation of this workflow to particular research communities, on an ad hoc, consultancy basis. The expertise that the BioExcel workflow development team has (arising from WP2 work) is quite extensive and therefore the aim is to capitalise on it.

The consortium understands that consultancy per se is not easily scalable as it depends on man power, however this is simply the first step to better understand the market interest in workflows and to serve as inputs for future developments. If enough interest is generated, the partners are committed to work more into the customisable manipulation of workflows through better user interface design in future development rounds.

Thus, we will start with promotion of the selected workflows to potential customers. We will offer training on how to modify and extend them on your own. We will also offer custom tailoring with BioExcel expert effort. If given workflows generate sufficient interest, we will explore the development and sale of GUI based solutions, which will offer an even smoother user experience, while being a scalable sale product in itself.

Action point(s)	Due
Definition of potential industrial users of workflows and organisation of focus group to gather feedback	September 2018

5.5 Other offerings

The BioExcel’s webinar series has shown to be popular among the communities with up to 100 attendees and up to 1000 views of the recorded presentations. Companies are interested to reach out to our community, and we have allowed one selected ISV to use the channel to showcase their products. Furthermore, several pharmaceutical companies have requested site visits by our lead experts in GROMACS and HADDOCK for in-depth consultation on usage of these codes and scientific aspects of the applied techniques. However, we do not expect that access to the webinar series or sites visits are suited to generate significant revenues. Site visit require the participation of PIs, and are thus difficult to scale. And the webinar series would lose its attraction for the largely academic community, if it were extensively used to promote commercial software. Rather, we see access to the webinar and site visits as confirmation of interest from industry in our expertise, suited to attract industrial customers to our consultancy or training offerings.

One such example is our participation in a joint proposal for Integrated Modelling of Antibodies (IMAb) in collaboration with three large Pharma companies. Several BioExcel partners participate in this proposal with activities that are directly aligned with the BioExcel core activities, namely development and application of BioExcel-supported codes towards understanding the biomolecular dynamic aspects of antigen and antibody interaction, and application of tailored competency profiles for training of computational researchers. Whether the proposal is successful or not, the writing process has established a strong group of partners who will continue to work together on BioExcel-related activities.

6 Marketing Strategy

As stated initially, we have the goal to build and grow an active user community, but we also need to gain the attention of potential customers for our commercial offerings. Marketing and branding are part of core activities to achieve this goal.

1. Gather access points	- Combine support forums - Provide downloads & manuals through one website
2. Provision of support	- Support forum, knowledge resource center - Tutorials, webinars, training material

3. Branding	<ul style="list-style-type: none"> - Clean and easy to navigate webpage - Same look and feel in all communications
4. Marketing	<ul style="list-style-type: none"> - Success stories, videos, brochures & blog postings - Social medial and direct reach-out

6.1 Branding

The BioExcel brand is starting to be recognised in biomolecular research circles, especially in the countries where we have organised events during the project or where we had very active partners. A strong brand and a strong name are the basis of many of the other marketing efforts that we will do for the Centre. The BioExcel logo has been successfully used in all media. A tag line “Powered by BioExcel” can also become part of our branding image for the future and this could also be applied to some of the exploitable products (training, workflows). Similar, we intent to use a strong and simple motto such as "Connecting people, driving science" on landing page and flyers.

6.2 Online Marketing

Most of our exploitable results are technology products, and therefore it makes sense to push their promotion on online media. This has the advantage that people will be able to click and immediately view information about the product, but also is the simplest way to reach a target audience that is already quite technology aware. Online marketing is also more affordable than traditional media marketing, and gives a level of tracking that cannot be provided by the offline world. Online marketing activities will be split into several categories.

Social Media:

We have already developed a following on Twitter and established a LinkedIn site during the project lifetime. Our regular posts reach hundreds of people and boosting these on a regular interval will give us the traction we needed. BioExcel will therefore continue to push these channels by ensuring that interesting content is posted.

BioExcel.eu website:

The website serves as a marketing tool since through it, we can promote exploitable products (like the training, workflows, events etc.). In the future, we can give the portal a visual update to make it look more user-friendly and channel people to the right links. We could also look at multilingual presentations to reach industry contacts and SMEs in a variety of countries.

Online platforms:

Technology or industry-specific platforms are another media to advertise our services, such as Einfracentral.eu, which is a common place for promoting services by EU projects.

Webinars:

Regular webinars are another way of attracting interest in our centre and products. In general, our free webinar series is focused on explanation of our tools and research strategies. However, this series can also be used to promote some of our paid courses, or workflows to our community.

6.3 Offline Marketing

Part of our target audiences, especially the academics, research organisations, industry contacts and local SMEs might not be so reachable with online media, and therefore an intensive approach towards other forms of marketing needs to be done. Unfortunately, all offline channels are pretty expensive when compared to their online counterparts, so the choice on where and what to do needs to be carefully made. The following are some examples of channels that are being developed.

Networking with industry:

Our existing network of industry contacts is an asset both in business and in marketing terms. The organisations that have signed up with us in our community interest groups already know us, and therefore they are our first port of call to continue pushing our marketing message. Many are already interested to learn more about the continuation of BioExcel after end of project, and therefore we would be preaching to the converted. One-to-one business meetings, either online or in person where possible will help to cement our relationship with these multipliers. At the same time, we will continue looking for other industry contacts and organisations in order to re-enforce our network and also reach potential new markets.

Events:

Participation in biomolecular events is another important part of our marketing campaign. In all countries where we are present one can find different workshops, symposia, conferences and similar activities where industry, consultants, academia and SMEs meet in order to discuss the future of business. It will be vital for BioExcel CoE to continue being visible at such events. There are also specific events similar to the European Conference on Computational Biology (ECCB) Annual Conference where BioExcel was already heavily participating thanks to our dissemination partners. A momentum needs to be kept and therefore this should continue to be an annual trip. Where possible, the members of our Centre should attend their closest business events with the BioExcel.eu branding to push our products.

Articles in newspapers / magazines:

Another important offline medium is the multitude of newspapers and industry magazines that reach out to our target groups. Even though such traditional magazines or newspapers are not that popular anymore, they are still around, and are often regarded as a more trustworthy medium than online articles. Being featured in some of these magazines will increase the chances that potential clients will want to try out the BioExcel.eu tools and services. With a number of such articles, we can reach out to different journalists and bloggers that can refer to these printed media, to have them review our tools and to further spread the word.

6.4 Timeline

This plan depends on the budget available for marketing. However, even with minimal budgets we can get started, and improve our efforts as the centre grows. Posting on social media and email reach-out will be a regular activity in order to keep the online activity on the BioExcel.eu healthy and active.

Quarter	Actions	Target Client
---------	---------	---------------

Q4-2018	Protein Structure Determination Industry Meeting - November 11-13 – Webinar – launching Training	Innovative SMEs & Researchers Interest Group – existing
Q1-2019	Online Media Campaign – introducing new products/services Activities with industry – workshops in local areas	SMEs & Academia Industry
Q2-2019	Industry Conference - 20th Annual Drug Discovery Summit and 7th Annual Discovery Chemistry and Drug Design Congress Printed Media Campaign – focusing on new products/services	Industry & Researchers SMEs & Industry Consultants
Q3-2019	Webinar – focus on particular BioExcel tools Own event (workshop style)	New Industry contacts SMEs & Academia

6.5 Sales Strategy & Lead Generation

A pipeline of leads will be developed from the sustainability team in collaboration with all the partners and through the use of an online CRM (Customer Relationship Management) software like Pipedrive. We will strive to move these leads onwards and prepare them for commercial service adoption. We will be piloting training services such as the above described workshops or consultancy offerings, in order to market test them, and we will use these leads to check our assumptions with regards to pricing of services and willingness to pay. The aim is not to have thousands of leads, but to have a few good ones that we can progress with along the path to commercialisation.

The CRM software is currently being trialled by the WP5 team in order to see if it would be adequate for the needs of the consortium. Costs of access until the end of the project will be covered by the budget for user group engagement under partner UEDIN.

Action point	Due
Inventory of all relationships/contacts that BioExcel partners have with industrial actors. Likewise, inventory of all partnerships with non-commercial organisations. Integrate all this information into our CRM.	September 2018

7 IP Management

7.1 Description of software licenses

HADDOCK consists of a collection of Python, CNS, and other additional scripts and programs (csh, awk, perl, c++). It is developed by the Bonvin Lab at Utrecht University. The distribution can be obtained free of charge by completing a licence agreement form, stating that the software is used for non-commercial, scientific purposes only. Additional software and licenses required to run HADDOCK (e.g. CNS) need to be obtained directly from the distribution sites. Commercial users can buy a licence for local installations (free updates and some support are limited to one year) or for access to the web-server (starting fee + pay-per-use).

GROMACS is written in many C-dialects, but will successively transition into pure C++11 code. The software development currently receives contributions from developers at KTH (BioExcel), Uppsala University, and University of Colorado. The copyright holders have agreed to use the GNU Lesser General Public License (LGPL 2.1 or later), which, in summary, states that:

1. Any derivative of GROMACS must have a compatible licence, i.e. if distributed, the source code must be shared.
2. Any other software can be linked with GROMACS as a library, regardless whether it is free or proprietary.

The motivation for this choice was to minimise legal entanglements and to ensure that GROMACS stays free. Consequently, this permits to make modifications to the software and allows usage inside a company. But if a derivative work beyond linking is created that work has to be redistributed under the same terms.

PMX uses GROMACS as an engine to perform free-energy calculations. As such, PMX is a pre-processing tool, and a tool for analysis of the output. It automates the calculation of free energies for amino acids, nucleic acids, and ligands. PMX uses standard open-source Python modules distributed under different licenses, which all permit usage in commercial settings, and is itself published in L-GPL 3.0. Copyright holders are still very few, namely Daniel Seelinger, who started the development of PMX in 2006 and who now works at Boehringer Ingelheim, and more recently Bert de Groot (BioExcel) and Vytautas Gapsys (BioExcel). In contrast to the other software discussed above, the recent working version of PMX can be regarded as a purely BioExcel-financed software. PMX can be found at GitHub, and in 2017 a web-server version was made available to users. PMX has received positive reception among researchers in academia and industry for its good predictive results. In contrast to the available approaches that use equilibrium protocols, PMX is using non-equilibrium protocols for the calculation of the free-energy (all such protocols can be found in the public domain). Industrial users of the software are Boehringer Ingelheim (Daniel Seelinger) and Janssen (BioExcel use case). The modification of ligands or mutations is currently performed with a script (semi-automatic), but the intention is to develop a fully automated workflow.

Workflow components and global workflows developed during this phase of BioExcel are expressed in Common Workflow Language (CWL), and Python is used to build wrappers that are needed to access applications within a workflow. Similar to GROMACS and HADDOCK, these workflows received contributions from outside of BioExcel. The CWL developers have decided to apply the Apache License 2.0 to all workflow components. In particular, this license does not require collection of IP transfer for accepting third-party contributions, and permits development of derived black-box commercial software, as long as the copyrights are attributed; meaning that the BioExcel workflow components can be used as starting points for commercial workflows.

CPMD is licensed by IBM. After registering on the CPMD website, the can be used for academic and non-commercial purposes. For commercial purposes direct agreements with IBM are needed.

Within BioExcel, we have developed the external MiMiC library for CPMD, licensed under LGPL v2.1, in order to enable CPMD to perform hybrid QM/MM simulations in combination with the GROMACS code. The MiMiC project have required some minor modifications in both the CPMD and GROMACS codes that will be available in the next releases of these codes.

7.2 Commercial aspects

The software and the workflows whose development is supported through BioExcel can be used in commercial settings, except for MiMiC that depends on CPMD. The open-source codes are published and covered by licenses that ensure freedom-to-operate. Regulatory aspects that commercial users of such software would need to observe within different industries are not considered at this stage.

Action point(s)	Due
Definition of IP for the training content.	September 2018

8 Legal Structure

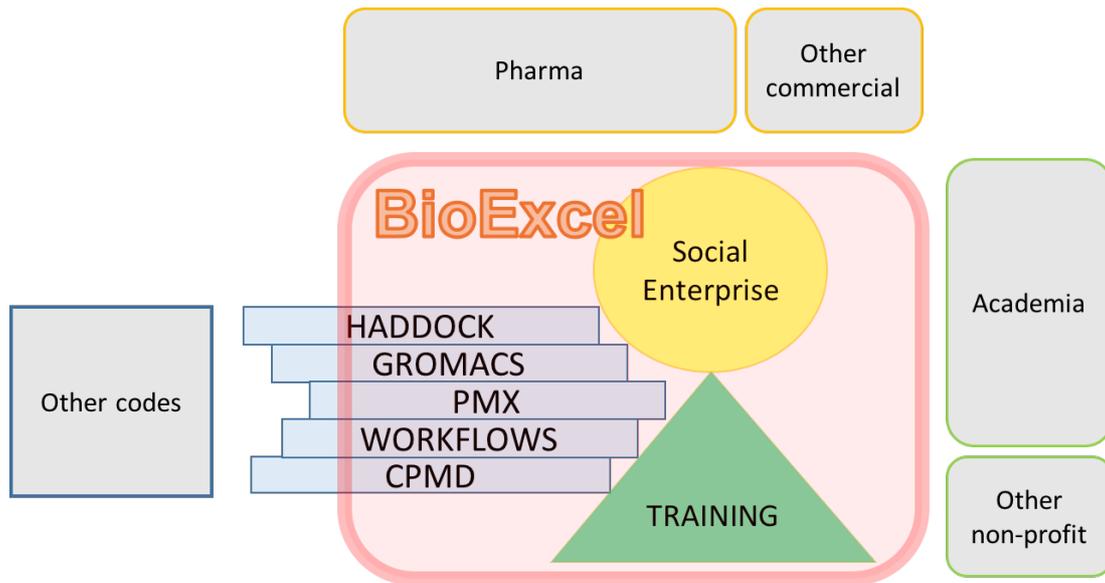
BioExcel will consist of two structures that co-exist to assume different functions. This setting will provide flexibility and efficiency, while keeping the cost at a minimum.

The overarching structure is the consortium of software developers and providers of user support and training, foremost aimed to benefit non-profit users. This consortium is only in place when public funding is provided. But even if no public funding is provided to BioExcel, most of the codes have some other sources of public funding available that finances some degree of operation, maintenance and development. This is visualized by their extrusion through the BioExcel framework in the figure below.

A limited company will be established, available to all consortium partners for commercial interactions, foremost with industrial users, but also with non-profit users. In other words, this company is BioExcel’s single interface for commercial activities. The company can and will also operate when no public funding is provided, and will be set up as social enterprise. As such it will endeavour to cover all its operational costs - direct and indirect from the revenues that it generates. If there is a surplus, this will be kept as a reserve to be used and reinvested in the top priorities decided by BioExcel management. These could potentially include:

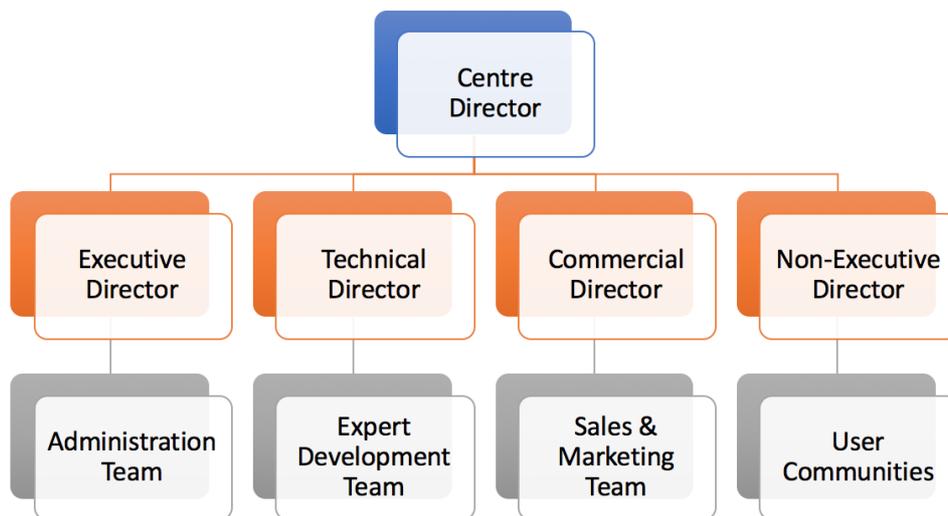
- Investment in networking meetings between the CoE and its communities
- Software development not currently being covered by public funds
- Promotion of the BioExcel CoE

Consortium and company will bear the same brand. Any BioExcel-generated IP that is intended for commercial use, as well as trade knowledge, shall be protected under the roof of this social company.



8.1 Operations Management

The day to day operations of the commercial legal entity that will be instrumental for the sustainability of BioExcel will require a special coordination team that will remain operational even after project end. Having lean operations is crucial in the embryonic stages of development of a commercial legal entity - even if this is a social enterprise. Moreover, it is important that there is continuity from the project phase to the post-project phase. The governance will take the user perspective into account through inclusion of a non-executive director. At the moment, the different roles envisaged are the following:



8.2 Modus Operandi

The legal entity will strive to operate as a channel through which the expertise and the team that together makes the BioExcel Centre of Excellence can interact with the external world. This is especially important since it will offer a simplification of the

commercial type of interactions like invoicing and receiving of funds that might otherwise be impossible for some partners.

The new entity will therefore be the one that will be engaging with the commercial actors, sending proposals, running promotional events and where possible marketing itself as the voice of the computational biomolecular expert community. It will act as an intermediary for the sourcing of expertise and create a “one-stop shop” for industry that wants to collaborate with the BioExcel CoE.

The entity will be able to enter into contracts for the provision of services towards third parties, and then it will have collaboration agreements with the different BioExcel partners that will actually provide the services.

Action point(s)	Due
Naming the first set of management personnel and team	November 2018
Draft collaboration agreements with BioExcel partners	December 2018 and beyond

9 Financial Management

Every organisation, be it profit or non-profit needs to manage its finances and the continuation of BioExcel will be no exception. If we are looking at BioExcel as a whole, which is made up of the partners that have been brought together through the Horizon 2020 project, and also the nascent legal entity, we can foresee that there will be three layers of funding:

- **Funding directly aimed at individual partners** - This will continue to support the development of the codes in their native form. Funding could be financial or in-kind, by donating development time towards improvement of the different codes, independent from BioExcel.
- **BioExcel public funding** - This include regional, national and supra-national grants that the different partners are able to attract in order to continue their collaboration and to focus on particular areas. For example, H2020 grants for research but also ITN or COST Actions.
- **Commercial funds** - This includes all the different revenue generating activities that the social enterprise will be engaging in. Income could be from direct sales of products or services to clients or via investors.

9.1 Social Enterprise Startup Funds

The costs of registering and starting up the legal entity will be dealt with by the end of the project timeline in order to make sure that it is possible to bootstrap the entity and give the required push.

Once the entity is registered, based on its jurisdiction, the management team will apply for start-up funds under local innovation entities in order to get an initial assistance. For example, if the entity is registered in Sweden, the national funding agency VINNOVA provides assistance to innovative start-ups on an open call policy.

9.2 Social Enterprise Operational Funds

Daily operations will require financial management and covering for the following types of costs:

Fixed Costs:

- Legal and Auditing Fees
- Administration staff / management
- Banking charges

Direct costs

- Marketing and PR costs
- Expert / Consultants / Trainers costs
- Travelling costs
- Infrastructure costs (physical / compute)

The aim is to have as much as possible a virtual organisation so that there are initially minimal or no costs related to rent, electricity, etc. With respect to staff members, even here, the idea is to keep the structure as lean as possible, and where needed to apply a “pay per use” type of employment contracts so that there are no or little fixed staff costs.

Needless to say, once the entity starts growing, then the management can go out to the investment community - business angels and venture capitalists in order to get a large injection that will allow the operations to grow. However, this will only happen at a later stage once the model has been proven to work.

9.3 Social Enterprise Revenue Management

Incomes from operations and sales will be managed by the entity and through direct invoicing of clients. At the moment, the expected types of income will be the following:

- Training Fees
- Consultancy Fees
- Income from ancillary activities such as site visits or access to the community

Action point(s)	Due
Budget projections based on different scenarios	November 2018
Achieve an understanding of costs associated with running BioExcel	November 2018

Appendix 1: Timeline and action points

The final business plan is due in project month 38, which is December 2018. Until then, we will further develop this business plan. The table below summarizes the actions points that were presented throughout the sections above.

Action point	Due
Update the table of literature references and the number of HADDOCK web portal users with fresh numbers from 2017.	June 2018
Refine the estimate of the number of industrial researchers by accessing publicly available data on industry sectors (e.g. CORDIS).	June 2018
Definition of potential industrial users of workflows and organisation of focus group to gather feedback.	September 2018
Inventory of all relationships/contacts that BioExcel partners have with industrial actors. Likewise, inventory of all partnerships with non-commercial organisations. Integrate all this information into our CRM.	September 2018
Definition of IP for the training content.	September 2018
Extend market analysis to understand opportunities for workflow-related offerings.	October 2018
Detailed definition of training product business model and competitive analysis to define a pricing strategy.	October 2018
Demonstrate what income can be generated from each activity, and discuss income targets.	November 2018
Naming the first set of management personnel and team.	November 2018
Budget projections based on different scenarios.	November 2018
Achieve an understanding of costs associated with running BioExcel.	November 2018
Draft collaboration agreements between BioExcel social enterprise and BioExcel partners that are involved in provision of commercial services.	December 2018 and beyond