

# H2020 EINFRA-5-2015



[www.bioexcel.eu](http://www.bioexcel.eu)

Project Number 675728

## **D5.2 – Revised business plan**

*WP5: Governance and Sustainability*



Copyright© 2015-2018 The partners of the BioExcel Consortium

## Document Information

<b>Deliverable Number</b>	D5.2
<b>Deliverable Name</b>	Revised business plan
<b>Due Date</b>	2017-04-30 (PM18)
<b>Deliverable Lead</b>	FT
<b>Authors</b>	Holger Noack (FT), Ian Harrow (IHC), Alexandre Bonvin (UU), Adam Carter (EPCC), Rossen Apostolov (KTH), Erwin Laure (KTH)
<b>Keywords</b>	Sustainability, users, support, services, cost and revenue
<b>WP</b>	WP5
<b>Nature</b>	Report
<b>Dissemination Level</b>	Public
<b>Final Version Date</b>	2017-04-29
<b>Reviewed by</b>	PMB
<b>MGT Board Approval</b>	2017-04-30

## Document History

<b>Partner</b>	<b>Date</b>	<b>Comments</b>	<b>Version</b>
FT	2017-04-17	First draft	0.1
FT	2017-04-24	Input from IHC, UU, KTH	0.2
FT	2017-04-26	Input from UU, KTH	0.3
FT	2017-04-27	Input from EPCC, UNIMAN, KTH, IHC	0.4
FT	2017-04-29	Input from KTH	0.5

## Executive Summary

This document is a revised business plan, based upon the previous deliverable that was submitted 6 months ago. It presents the current stage of findings and has to be viewed as intermediate step towards the final Governance and Business Plan that will be due 18 months from now.

The progress that BioExcel has made during the last six months has resulted in a refinement of our strategic development goals.

The vision and mission statements have undergone slight adjustments to emphasize BioExcel's focus on biomolecular simulations and modelling, and to highlight the desired impact of enabling better science.

While the previous deliverable D5.1 contained a more general description of the HPC ecosystem, we have focused our attention in this revised business plan on our anticipated user base.

The value proposition and the description of our support & service offerings have been refined. We acknowledge that we need to focus on where to generate value, and how we are going to do that, in order to match our ambitions with the available workforce during this initiation phase.

The section on financing has been adapted to reflect our change from the initially discussed membership-fee model towards a service-fee based model. Membership fees would have introduced exclusivity to our service offerings, which would have reduced the desired impact on the largely non-profit user base, and would have forced us to provide a parallel offering of basic (free) and premium grade services. The herein presented service-fee based model has a higher likelihood to be able of generating broad value to the biomolecular research community, while still providing us with the opportunity to collaborate with paying users and industry on single-standing projects.

The material in this deliverable is complementary to two other deliverables submitted at project month 18, as follows:

"D3.2 - Consultancy proposals" describes in what other projects the partners are involved. We argue that there are multiple funding streams for separate activities done by the partners, but in order to ensure a European leadership in the field, there is a crucial need for funding for coherent joint efforts.

"D3.3 - Consultancy Modalities and Funding Options" describes what services and in what ways (i.e. modalities) they can be offered by the center. We give examples of existing services provided by institutions at the partner countries.

This deliverable "D5.2 - Revised Business Plan" describes how those services will be financed and what legal structure will be able to support those offerings.

## Contents

<b>1</b>	<b>INTRODUCTION</b>	<b>6</b>
1.1	VISION AND MISSION	6
1.2	SUCCESS GOALS	6
<b>2</b>	<b>USERS</b>	<b>7</b>
2.1	USER SURVEY	8
2.2	REFERENCE USER BASE	9
2.2.1	ESTIMATING THE SIZE OF THE USER BASE FROM LITERATURE	10
2.2.2	ESTIMATING THE SIZE OF THE USER BASE FROM REGISTRATIONS	10
2.2.3	ESTIMATING THE USER BASE: CONCLUSIONS	11
2.3	ATTRACTING USERS	11
<b>3</b>	<b>VALUE PROPOSITION</b>	<b>12</b>
3.1	ADDED VALUE TO USERS	12
3.2	VALUE PROVIDED TO PARTNERS	13
3.3	VALUE PROVIDED TO OTHER STAKEHOLDERS	13
<b>4</b>	<b>SUPPORT, SERVICES AND TOOLS</b>	<b>14</b>
4.1	SUPPORT OFFERINGS	14
4.2	SERVICE OFFERINGS	15
4.3	TOOLS	17
<b>5</b>	<b>FINANCING</b>	<b>17</b>
5.1	STAFFING	19
5.2	WORKLOAD ESTIMATE	19
5.3	COST AND REVENUE	20
<b>6</b>	<b>LEGAL STRUCTURE</b>	<b>21</b>
<b>7</b>	<b>GOVERNANCE</b>	<b>23</b>
	<b>APPENDIX</b>	<b>24</b>
	<b>I. LIST OF ISVs</b>	<b>24</b>
	<b>II. NATIONAL FUNDING OPPORTUNITIES</b>	<b>24</b>

## 1 Introduction

For this business plan it is helpful to view the BioExcel CoE as an organisation under development. The current phase of organisational development (Phase 1) is earmarked by initiation of activities, verification of user needs, and testing of reception for proposed support and service offerings.

With verified user interest, support and service offerings can be formalized, the organisational framework can be implemented, and the center can commence full operation (Phase 2).

Thereafter, the center will be able to grow with its user base, adapting software development and service provision to changing user needs, and changing capacities of new hardware.

Initiation of software development	Definition of support and service offerings	Verification of user interest and impact	Installation of organisational framework	Full operation	Growth
Phase 1			Phase 2		Phase 3

### 1.1 Vision and mission

BioExcel aspires to become a central hub where biomolecular researchers, software developers, computational experts, and resource providers are linked together. Through BioExcel, exchange of knowledge and collaboration between these stakeholders shall be facilitated.

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

The impact goal of BioExcel is to promote (qualitatively) better biomolecular research, which also allows to address new and more complex research questions. This can be achieved by empowering users in how to use computational methods, and by providing users with improved computational tools of higher efficiency, usability and ability to scale with modern hardware.

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

### 1.2 Success goals

BioExcel has identified 13 long-term success goals that align with the mission and vision statements. These success goals help to generate our value proposition and to align the efforts of the BioExcel partners.

<b>USERS</b>	
1. Diverse user community	Attraction of users with different expertise and from different biomolecular disciplines.
2. Active user community	Users should actively contribute to help create attraction and visibility.
<b>SUPPORT</b>	
3. Expertise sharing	Facilitate sharing of expertise among users and experts, help with competence matching.
4. Collaboration	Encourage collaboration among users and with related organisations
<b>SERVICES</b>	
5. User training	Provision of training that enables users to make the most of software and workflows.
6. Consultation	Provision of various consultation offerings to work on defined, user-specific projects
<b>SOFTWARE</b>	
7. State-of-the-art software	Provision of easy-to-use, scalable, and efficient software with good documentation
8. Workflows components	Provision of tools that integrate the software into workflows, increase user-friendliness and efficiency
9. Integration of feedback	Facilitate integration of user feedback into software design cycles
<b>PARTNERS</b>	
10. Work satisfaction	Recognition of excellence for BioExcel staff
11. Career opportunities	Facilitate access to new research and development projects for BioExcel partners.
<b>ORGANISATION</b>	
12. Strong brand	Professional online presence and active marketing to improve visibility of our software and services
13. Stable financials	Diversify funding streams to become less dependent on punctual public funding

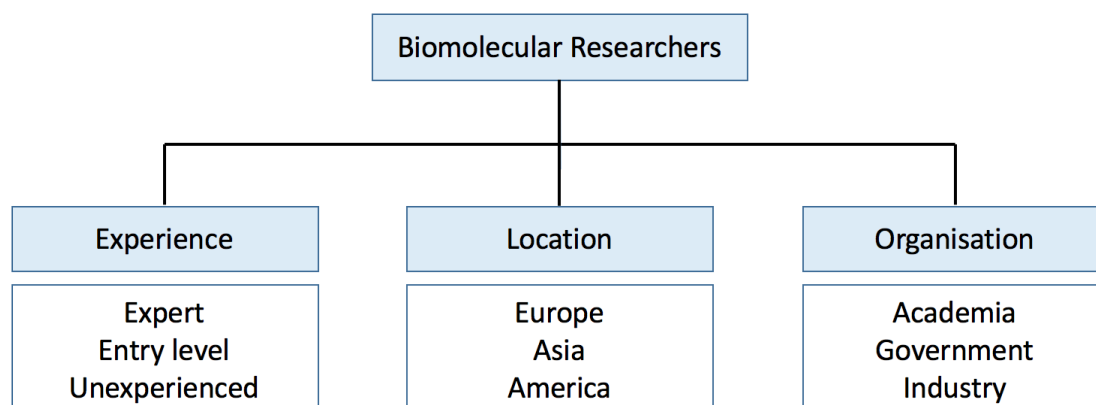
## 2 Users

The most important stakeholders in the value chain of biomolecular modelling and simulations are:

1. Biomolecular researchers
2. Software developers
3. System administrators
4. Hardware providers

BioExcel has the intention to work towards, or collaborate with all these stakeholders. However, during the current initiation phase we will focus our resources onto the end-users, i.e. the biomolecular researchers that use software for simulations and modelling. These prospective users can be segmented by their level of experience

with computational research, by their location, or by the type of organisation that they originate from.



Taking into account the level of user experience is especially important when designing workshops and other training material. User experience and gap analysis of user expertise are discussed in detail in work package 4 that deals with training and education.

The geographic location of users does not require further explanation. While support and software tools are non-exclusive, available to all users regardless of location, BioExcel will focus its free and publically-funded service offerings on users located in Europe. A detailed description of our service offerings can be found in section 4. However, where given we will collaborate with organisations internationally, such as the Molecular Science Software Institute (MolSSI) in the USA.

Differentiating researchers by their organisational types helps to understand user needs and motivation better. Academia has a broad support of post-graduate students and post-docs, who usually are open to use open-source software. Researchers from non-academic governmental organisations tend to work more towards applied research, and are thus likely to have higher requirements for stable tools that are easy to use. Researchers from industry that work with product development require software with version control to validate production of a data packages. Those researchers are less likely to use open-source software, or will do so only in early stages, where regulatory restrictions do not apply.

## 2.1 User survey

The table below provides some insight into the user communities from GROMACS and HADDOCK that were obtained through surveys. It can be seen that the overwhelming majority of users are non-profit/academic researchers, with a high percentage of PhD students, and a broad global spread.



Survey findings	GROMACS <sup>1</sup> 224 responses	HADDOCK <sup>2</sup> 697 responses
User type	94% non-profit 6% for-profit	97% academic research 3% industrial research
User location	50% Europe 19% North America 10% Rest of America 16% Asia	28% Europe 19% North America 24% India 14% Rest of Asia
User level	43% PhD 22% Postdoc 14% Staff scientist	34% PhD 32% Staff scientist 20% Postdoc
User experience	Not asked	34% use code for 1-3 years 42% use code for < 1 year 24% use code for > 3 years
Research interest (2 answers possible)	27% Chemistry 20% Molecular Biology 17% Bioinformatics	75% protein/protein 32% protein/peptide 23% protein/DNA
Publication of results	Not asked	57% no 43% yes
User satisfaction 1-5 = low-high	Not asked	51% grade 4/5 31% grade 5/5 14% grade 3/5
Type of installation	Local cluster > Supercomputer > Workstation > Cloud	91% web server 35% local

## 2.2 Reference user base

The objective of this section is to get an understanding of the size of BioExcel's user community, i.e. an estimate of the total number of users that we can reach with our value proposition.

Initially, BioExcel's users base will consist of the already existing users of the three software codes for which BioExcel is improving usability and performance. The provision of services and support will reach these users first.

Where users are required to register, the size of the user base can be estimated by counting the registrations. Another way to estimate the number of users is by counting the publications that refer to the biomolecular software used. However, when estimation the size of the community from citations, one needs to keep in mind that the software can be used in other scientific disciplines too. HADDOCK is solely used for biomolecular research, whereas GROMACS and CPMD also have an audience in the wider field of chemistry and physics.

<sup>1</sup> GROMACS feature and usage survey, 2014

<sup>2</sup> online survey, access on April 2017:

<https://docs.google.com/forms/d/1WMzzvssuDMApHJW8ugMODXeo34tCI6CmXcdJsRctTO8/viewanalytics#start=publishanalytics>

### 2.2.1 Estimating the size of the user base from literature

The table below shows the number of articles that were returned in a Google Scholar search, either citing or mentioning one of the BioExcel codes. In 2016, there were in total 4759 publications that cited one of the three codes.

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

### 2.2.2 Estimating the size of the user base from registrations

Since HADDOCK requires users to register, we are able to obtain a more detailed picture of our prospective user base coming from HADDOCK. Over the past five years, the number of registered users on the HADDOCK web portal has constantly increased (see figure below). In April 2017, the total number of registered users exceeded 8900<sup>3</sup>. 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.

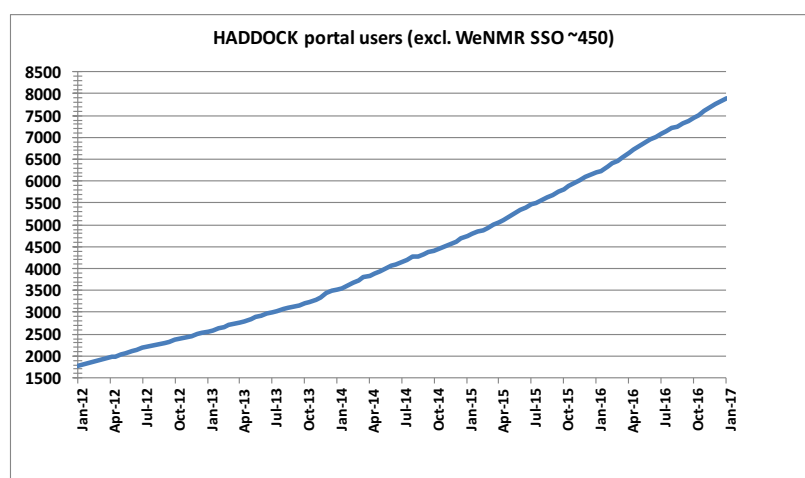


Fig: Number of registered users on the HADDOCK web portal between January 2012 and January 2017.

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

Not all of the registered users are at all times actively submitting docking runs. The table below shows the number of new registrations, and the number of active users during the last two years.

<b>Web portal</b>	<b>2015</b>	<b>2016</b>	<b>increase</b>
# 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 <sup>4</sup>	24806	28402	14%

Table: Breakdown of HADDOCK web portal use for the years 2015 and 2016.

For 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 say that for HADDOCK each publication represents on average 3 active users.

It is also worth noting that the server is also being used for teaching since we do observe registrations from students of some given university at specific times during the year. Those users will clearly not lead directly to publications.

### 2.2.3 Estimating the user base: conclusions

Through the survey, we know that at least 50% of the GROMACS users work in the field of biomolecular research, and based on the mailing list traffic a more realistic number is likely closer to 90%. Using the conservative estimate, we can assume that at least 2000 of the GROMACS citations in 2016 were related to biomolecular research. Adding to that the citations from HADDOCK and CPMD gives a total of roughly 2700 publications. In comparison, the more generous definition gives a total of 5000 life science publications that BioExcel-funded software contributes to.

If we then generalize the findings from HADDOCK, i.e. if we assume that each publication stands for 3 active users, we arrive at a biomolecular user base of circa 8000 to 15000 researchers. This should be seen as the minimal size of our user base, since we were limiting our estimation on the already existing users of the three software packages that currently are part of BioExcel.

## 2.3 Attracting users

Attracting users has been discussed during a workshop on sustainability with other CoEs and initiatives similar to BioExcel. The four activities below have been identified as important to facilitate attraction:

1. Gathering of access points (to unify disperse communities)
2. Provision of support
3. Branding
4. Advertisement

---

<sup>4</sup> The 24806 submissions in 2015 resulted in over 8 million HTC jobs.

The table below summarizes how BioExcel intends to work on those four activities to attract users. The last column indicates the phase when certain activities should take place. Phase 1 refers to the current project execution phase (2015 – 2018), and phase 2 is the anticipated next execution phase (2018 onwards). This separation is needed to match the workload with the available resources.

	How	When
<b>Gather access points</b>	1. Combine support forums	Phase 1
	2. Collect manuals & tutorials on one website	Phase 1
	3. Repeat 1. and 2. for new partner codes	Phase 2
<b>Provision of support</b>	1. Supervision of support forum & helpdesk	Phase 1
	2. Provision of webinars	Phase 1
	3. Provision of recorded workshop lectures	Phase 1
	4. Provision of knowledge base	Phase 1
<b>Branding</b>	1. User-friendly webpage	Phase 1
	2. Corporate identity (e.g. business cards)	Phase 1
	3. Same look and feel for all tools on web portal	Phase 2
<b>Advertisement</b>	1. Movie clips, success stories	Phase 1
	2. Brochures & blog postings (e.g. on LinkedIn)	Phase 1

### 3 Value proposition

It is BioExcel's intention to focus value creation on the end-users, i.e. biomolecular researchers working with computational simulations and modelling. Other stakeholders, such as software providers or administrators will be included during a later phase of organisational development, once the main user base has been established and the CoE has gained momentum.

To begin with, the three software packages that currently are part of BioExcel each already have an existing user community. To achieve sustainability, BioExcel would need to bring added value to these existing user communities.

Furthermore, the expertise offered by the CoE must be able to match the needs of the research community. To achieve this, BioExcel is required to also generate value internally to its partners, to the extent that the competence needed to serve the biomolecular research communities stays within the CoE.

#### 3.1 Added value to users

Prior BioExcel, software development was sporadic due to intermittent funding streams. It was not possible to provide reliable user support, to update software with evolving hardware possibilities, or to develop new features that meet changing user needs. It is thus the intention of BioExcel to provide added value to biomolecular researchers through:

1. Increasing user support
  - a. Provision of general support (forum, tutorials, knowledge base)
  - b. Provision of services & consultancy (project & user specific)
  - c. Provision of training and education (webinars, workshops)

2. Increasing usability of software
  - a. Provision of workflow components and containers
  - b. Development of new software features
3. Increasing performance of software
  - a. Improving scalability on modern hardware infrastructures
  - b. Improving efficiency of hardware use

A more detailed description how this added value will be provided is given in section 4 below. Through a fully functional BioExcel CoE, as envisaged in phase 2, we will be able to substantially improve user experience. The current funding levels during our initiation phase, however, do not provide sufficient resources to deliver all of the added value identified above.

### **3.2 Value provided to partners**

Collaboration partners that work with software development are central to the expertise of the CoE. On the one hand, these experts are needed to improve codes that are important to the biomolecular research community. On the other hand, they are also needed to support BioExcel's service offerings and consultancy projects with their expertise. For these partners, BioExcel could provide value by:

1. Financing of software development
  - a. Active search of public funding opportunities, coordination of application process, and attraction of funding to finance software development that increases usability and performance.
2. Reduction of time consuming tasks
  - a. First line in user contact through a helpdesk mechanism to manage user queries (competence matching)
  - b. Dissemination activities to increase user attraction
  - c. Organisation of hackathons
3. Support with debugging and packaging of code
  - a. Provision of a compile farm and hands-on support with updating containers

In its current stage, BioExcel is able to provide 1. and 2. from above. Provision of support for debugging and packaging of code was highlighted as useful addition by collaboration partners, and could be implemented in a later phase.

### **3.3 Value provided to other stakeholders**

As mentioned before, during this initiation phase BioExcel will focus on the needs of biomolecular researchers working with simulations and modelling. To broaden the impact, however, BioExcel will strive to also support other stakeholders of the value chain, such as administrators or independent software vendors.

#### **Administrators**

System administrators operate the hardware, and are involved during installation and performance tuning of biomolecular software. The quality of the software is part of the user satisfaction concerning the services provided, and also impacts the efficiency on how the infrastructure is used. BioExcel will work closely with system administrators, and support them in deployment and testing of software and updates.

### Independent software vendors

Independent software vendors (ISVs) are also working towards the biomolecular research user. The majority of ISVs are small or medium-sized enterprises that also must deal with bug fixes, development of new features, and support for a combination of operating systems, distributions, processors, and middleware environments. By implementing a policy of open and non-restrictive programming interfaces and support libraries, BioExcel would ensure that software created within the open-source community allows for ISVs to provide added value, i.e. to develop complementary commercial variants and related software tools. A list of currently 38 ISVs is attached in the appendix. The intention is to approach these ISVs during the next months to gather feedback and test their interest in collaboration with BioExcel.

## 4 Support, services and tools

As discussed above, value will be generated through provision of improved and new support offerings, services and tools. The differentiation between these is given in the table below.

Definition used	Characteristics
<b>Support</b> is an ongoing offering to users that is non-specific and non-exclusive.	Workload is relatively constant and predictable, and benefits an unlimited number of users.
<b>Services</b> are an ongoing offering that require active engagement with one user or a group of users.	Varying workload that requires flexible workforce, each point of service is a solution of a particular problem and benefits a limited number of users.
<b>Tools</b> can be obtained by downloading, or can be provided through a platform, and can be used without input from BioExcel.	Substantial workload needed for development of new, or improvement of existing tools (software/components). Once developed, a tool can be used by many users repeatedly.

### 4.1 Support offerings

The support offerings require a relatively stable workload and can be initiated already during the current phase of BioExcel.

Support to biomolecular researchers	Conditions	Phase
Website	None	1
Tutorials	None	1
Knowledge base	None	1

**Website:** All support offerings are made available through the BioExcel website, which will serve as a portal. Users will also be able to access information about service offerings, software tools, and relevant events.

**Tutorials:** Tutorials will be provided in the form of use cases and “how-to” descriptions (that might include video). On the long run the intention is to do that not only for the current three BioExcel codes, but also for any future code that might be taken care of through BioExcel. This aligns with the idea to bundle access points (user attraction) and unify presentation (branding, user attraction), which in turn will make it easier for biomolecular researchers to move between different codes.

**Knowledge base:** Users will be able to search the database for training resources based on topic, free text or event type. This concerns all training resources that are relevant to the biomolecular research community, i.e. not only BioExcel events. The aim is to integrate the BioExcel competence profiling into the database, which should help users to find training opportunities that match their personal needs. This work is described in more detail in work package 4.

## 4.2 Service offerings

We can categorize the services that we provide into three groups that relate to the user need and corresponding expertise needed:

1. Scientific services that deal with how to use biomolecular software to solve specific research questions
2. Technical services that deal with performance of our software, and software-hardware interplay
3. Training and education

A deeper analysis of the consultancy modalities of our services is provided in deliverable D3.3. The table below provides an overview over our services, indicating conditions and organisational phase during which we expect to be able to provide that service.

Services to biomolecular researchers	Conditions	Phase
<b>Scientific Service</b>		
Expert support on forum	Free, registration	1
Public drop in sessions	Free	1
Consultation meetings	Free, competence matching	1
Consultancy, short & medium projects	Fee, competitive peer review	1
Consultancy, long-term projects	Free, competitive peer review	2
Part writing of joint proposals	Free	2
<b>Technical Service</b>		
Expert support on forum	Free, registration	1
Performance validation	Free	2
Software installation	Free for non-profit	2
Feature development	Fee	2
<b>Training</b>		
Webinars	Free	1

Standard workshops	Free	1
Custom workshops	Fee	2

**Support forum:** BioExcel will monitor a user forum and provide input to the users, both on scientific and technical questions ([ask-bioexcel.eu](http://ask-bioexcel.eu)).

**Drop in sessions:** At some of our partner universities, local BioExcel experts will be available for 1-2 hours each fortnight for users to drop in and discuss all types of questions related to the use of biomolecular software for simulations and modelling.

**Consultation meetings:** These meetings have to be booked in advance, so that we are able to match a specific user need with the right competence. These meetings could be used to help users with project design.

**Project consultancy (short & medium term):** Project applications are reviewed every second week, and are selected based on available expertise and feasibility. BioExcel will charge a small fee (below self-cost) for consultancy on user specific projects. Time and outcome estimations are done individually for each project before contracting. The fee helps to define projects through agreement of clear start and end-points. The use cases currently performed under work package 3 are preparation for this service offering.

**Project consultancy (long-term):** Project applications are reviewed, and support is granted to a limited set of scientifically outstanding projects. Project are financed through public funding.

**Proposal writing:** BioExcel will help collaboration partners and users with preparation and compilation of joint-proposals to fund software development or long-term project consultancy. This service is also important to facilitate collaboration with SMEs and smaller ISVs that lack the funds to pay us full service fees. In this case BioExcel could help these industrial actors with joint-proposals to innovation agencies, for projects that might deal with e.g. scientific consultancy, customization of software, or co-development with ISVs.

**Performance validation & installation of software:** These technical services are mostly directed at system administrators.

**Feature development:** This service offering is provided with industrial users in mind, such as ISVs or large pharma. On request, we would be able to develop features or customize the software to work in a user-specific setting. BioExcel will charge a full fee on these type of user-specific projects. Time and outcome estimations are done individually for each project before contracting.

**Webinars:** Webinars are held regularly, and are published afterwards. A roadmap of upcoming seminars will be made available. Speakers from the user community and from partner organisations will be invited for some webinars.

**Workshops:** The majority of workshops will be provided free of charge and deals with general study design and use of software and workflows. In cases where substantial preparation and attendance of deep experts is required, a small fee will be



charged. A survey with HADDOCK users showed that 75% would accept to pay a small fee if course material and on-site support from developers is included.

### 4.3 Tools

Tools to biomolecular researchers	Conditions	Phase
Workflows	Free	1
Web portal	Free for non-profit	1
Software	Free	1

**Workflows:** Workflows will allow to integrate different software or automate their use, thereby increasing usability and efficiency of provided software.

**Web portal:** BioExcel will provide workflows and software through a web portal. The web portal removes the burden of installation and setup, while providing additional functionality through provision of validation routines for input data. For high throughput calculations it might be possible to provide web portal users access to EGI resources, as currently done by HADDOCK.

**Software:** The lack of support and maintenance has been the biggest issue for open source software. BioExcel will work with software development to improve scalability and functionality. During its initiation phase, BioExcel is supporting the development of three widely used codes, which taken together provide a powerful toolbox for computational, biomolecular research:

#### 1. Atomic quantum level: CPMD

The Car-Parrinello Molecular Dynamics package (CPMD) is an implementation of density functional theory, and is used to study smaller systems and their electronic states. In the context of biomolecular research it could be used to research polarization states of amino acids, proton transfers, or the effects of solvation.

#### 2. Molecular dynamics and 3D-structure: GROMACS

GROningen MACHine for Chemical Simulations (GROMACS) is based on classical molecular dynamics, and mainly designed for simulation of proteins, lipids and nucleic acids. It is very popular among users for being able to simulate extremely large systems considerably faster than other simulation software.

#### 3. Macromolecular assemblies and interactions: HADDOCK

High Ambiguity Driven biomolecular DOCKing (HADDOCK) software is driven by experimental information and enables the study of tertiary and quaternary structures of proteins and their interactions with other proteins or molecules.

## 5 Financing

BioExcel is built around open-source software with proven high academic and industrial impact that is freely available. Licensing fees would constitute a significant barrier to non-profit researchers, the overwhelming majority of the biomolecular

research community, and would hence impede any efforts to generate value. The same is true if our service offerings would be exclusive to paying members only. It is thus clear that the main source of financing for software development and provision of services and support is public funding.

However, to become less depended on punctual EU funding, a diversification of funding streams needs to be achieved. Possible sources are listed below, in order of their attractiveness for broad generation of value to the user community:

<b>Funding sources</b>
<b>Public funding</b>
1. European public funding
2. National public funding for single standing projects. Applications could be related to development of software towards specific user groups, or to collaborative research and development projects. A list of available funding opportunities on national level is provided in the appendix, and examples of funding streams that are already being used by BioExcel partners are discussed in more detail in D3.2.
<b>In-kind contributions</b>
3. In-kind contributions from partner universities
4. In-kind contributions from other stakeholders that benefit from effective and scalable software. These could be hardware providers interested in that codes relevant to the community perform well on their hardware (e.g. Intel).
<b>Fees</b>
5. Small service fees to non-profit users for some consultancy projects and selected training events.
6. Full service fees to industrial users for consultancy projects and training events.
7. Membership fees, which however would require BioExcel to bundle services and access to software into basic and premium offerings.

In this setting, industry funding will be attracted through consultancy projects, such as feature development or scientific consultancy as described in section 4. A more elaborate value proposition towards industry has to be developed during the next months to identify other means of attracting industry funding. We anticipate, however, that a large user community linked to our center will be important to gain the attention of industry, since it will serve as a reference for our ability to generate value, and as a pool for potential collaboration with industry.

Tapping into the public funding opportunities on national level, provided mostly from governmental organisations and foundations, requires dedicated staff to coordinate these efforts. The examples given in the appendix show that there is a diversity of opportunities, each with different deadlines, for different project topics, different impact goals, varying requirements on application partners. However, especially grants for innovation and valorisation activities provide a good opportunity to broaden our impact beyond academia. Our intention is to visit fairs and talk to innovation

agencies so that industrial actors are aware of our availability for joint collaboration projects.

## 5.1 Staffing

The cost of the center is mainly related to the cost of staff. We suggest three layers of staffing to engage broadly with the biomolecular research communities.

Staffing layers
<p><b>1. Software development</b> This staff category consists of experts working with development or improvement of specific codes that are taken care of through BioExcel. It will be up to the groups involved to show that their work (gradually) achieves large impact, and that the code is licensed on free/business-friendly terms.</p>
<p><b>2. Support &amp; Services</b> Staff working in this category needs to have some programming expertise and good experience from research in the field. Their role should be to engage with the entire community and help realize important challenges, but also to identify important parts that require development and improvement of code.</p>
<p><b>3. Management &amp; Outreach</b> Staff in this category will be responsible for managing the center's operations, dissemination, marketing &amp; branding, communication with industry and other stakeholders, but also coordination of activities and intra-organisational communication to ensure equal level of information among the BioExcel partners about the challenges of the biomolecular research community.</p>

## 5.2 Workload estimate

The workload estimate is provided for the three phases of organisational growth presented in section 1.

### Phase 1 (current phase)

We start by assigning the workforce of our current six work packages to match the three layers of staffing described above. Work packages 1 & 2 belong to the staffing section of software development, work packages 3 & 4 belong to support and services, and finally work packages 5 & 6 are counted into the staffing section for management and outreach. Based on this segmentation we come to following conclusion:

1. Software development: The available workforce for software development during phase 1 is circa 1.5 FTE per software package, plus another 1.5 FTE for development of workflow components and interfaces, i.e. a total of 6 FTE.
2. Support & Service: The current FTE workforce in this category is 5 FTE
3. Management Outreach: The available FTE in this staff section is 2 FTE

### Phase 2 (operational phase)

1. Software development: To achieve the desired progress we would need at least 3 FTE per software package and another 3 FTE for development of workflow components and interfaces, i.e. a total of 12 FTE.
2. Support & Service: From interviews with similar organisations we have learned that the provision of user support, training and consultancy requires

approximately 1 FTE per 20 user contacts (averaged through all the service types). Assuming 200 direct user-service interactions, e.g. in the form of consultancy projects or workshops, would require us to have 10 FTE in this staffing section. The number 200 corresponds to 2.5% of our estimated reference user base.

3. Management Outreach: Management and outreach is estimated to require 3 FTE for an organisation with international operations.

### **Phase 3 (growth)**

1. Software development: From a dialogue with developers, we estimate that the workforce for software development ideally should consist of 5 FTE for each code, as well as additional 5 FTE working with overarching workflow components and interfaces, i.e. a total of 20 FTE, if we continue to work on 3 software codes in parallel.
2. Support & Service: Assuming 400 direct user contacts throughout our service portfolio and related support offerings requires 20 FTE in this staffing section.
3. Management Outreach: Management and outreach is estimated to require 5 FTE with this larger organisation, mostly due to an increasing workload to attract funding from national grants and industrial actors

## **5.3 Cost and revenue**

### **Estimation of cost**

The cost and revenue estimate is based on the three staffing categories and the estimated workload during the different organisational phases presented above. The average cost of salary in BioExcel is €7000 per person month. Factoring in 60% overhead, which is common for universities, increases the cost per FTE to approximately €11.000 per month.

### **Estimation of in-kind contributions**

For phase 2 and phase 3 we try to achieve in-kind contribution from partner universities for the staff working with management & outreach (aim for 20% in-kind), as well as the staff working with support & services (aim for 10% in-kind). Similarly, we try to achieve in-kind contributions from hardware providers for the staff working with software development (aim for 10%). The motivation for hardware providers is to ensure that codes important to the community run well on their hardware.

### **Estimation of service fees**

As described in the service section, we will charge a small fee to non-profit users for small and medium scale consultancy projects. To stay affordable, we will probably charge a fee corresponding to 20% of our FTE cost, corresponding to circa 10-15€ per hour. For industrial users, we will charge our full FTE cost. The reasoning is similar for those of our workshops, where a fee will apply. However, the consultancy projects and the workshops that generate these fees are only one part of our otherwise free service offerings. Hence, we estimate that these fee-generating activities contribute to a total of circa 5% of our FTE cost for the support and service staffing.

The estimated cost and revenue streams for our three phases of organisational development are summarized in the table below.

Cost & Revenue Estimate	Phase 1	Phase 2	Phase 3
Software development [FTE]	6	12	20
Support & Services [FTE]	5	10	20
Management & Outreach [FTE]	2	3	5
TOTAL FTE	<b>13</b>	<b>25</b>	<b>45</b>
Annual cost, circa [€]	1.7 million	3.3 million	5.9 million
In-kind [€]	-	0.37 million	0.66 million
Service fees [€]	-	0.07 million	0.13 million
Public funding [€]	1.7 million	2.9 million	5.1 million

## 6 Legal structure

The services that we intent to provide require us to deal with short-term projects, e.g. consultancy or training. Here, we cannot rely on our partner universities to provide us with their experts whenever needed. Instead, we need to be able to independently employ personnel that is available to answer varying project demand.

Furthermore, as described above, we intent to charge service fees for some consultancy projects and training events. These intentions must not necessarily overlap with the operational capabilities of our partner universities. In addition, users and BioExcel experts maybe located in different European countries. The goal is thus to become an independent entity that is capable of making independent strategic decisions, enter into contracts, and perform financial transactions across borders. The most important requirements on this legal entity are summarize below.

<b>Requirements on future legal structure</b>
Enable value creation
Possibility to make independent strategic decisions
Possibility to enter into contracts
Possibility to be employer
Eligible to receive grants
Possibility to charge service fees
Possibility to operate across borders
Ownership of brand

We have investigated organisations similar to ours, who have transitioned from public funding to a sustainable model such as OpenPhacts or FAIRDOM. The finding was

that most service providing organisations are set up as non-profit, i.e. as foundations or charities. Another legal structure that seems to be suitable for some academic consortia is the European Economic Interest Group (EEIG). The table below shows how the most popular legal structures compare on requirements that are important to BioExcel.

	<b>Collaboration agreements between universities</b>	<b>Foundation</b>	<b>EEIG</b>	<b>Company</b>
Access to EU funding with research focus	Yes	Yes	Yes	No
Access to EU funding with innovation focus	Limited	Limited	Yes	Yes
Membership fees	Yes	Yes	Yes	Yes
Commercial service fees	No	No	Yes	Yes
Freed from charging VAT	No	Yes	No	No
Freed from cooperate tax	Yes	Yes	Yes	No
Can be employer	No	Yes	Yes	Yes
Independent decision making	No	Yes	Yes	Yes

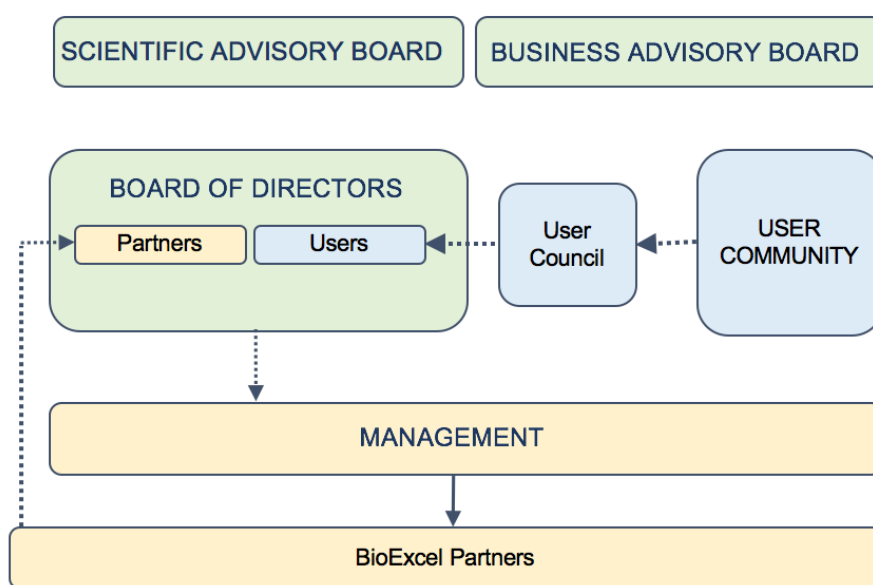
Foundations or charities are ideally suited for organisations that in addition to public funding intent to finance their operations through membership fees, since they are not required add VAT on those membership fees.

Membership fees have been discussed initially among the BioExcel partners. However, during the refinement of our value proposition it became clear that membership fees would constitute a significant barrier to generate value for the larger biomolecular research community. Either our entire offering to the community would need to become exclusive, which would go against the philosophy of open and free access, or we would need to provide two parallel support and service offerings of different quality, to provide added benefit to users paying membership fees.

We have therefore decided to keep the majority of our offering free, and instead charge service fees for some of our consultancy and training offerings, as presented in section 4 and also in deliverable D3.3. At current, it is thus our belief that an EEIG would be a suitable legal structure to house or operations. An EEIG would be able to take part in EU programmes, while also being able to interact with industry and employing its own personal. Lately some universities have started to join EEIGs, so we can expect pre-existing knowledge in some academic, legal departments on how to proceed. The intention is to start the dialogue with all partner universities during Q3 2017, so that a draft agreement is in place ideally by the end of phase 1.

## 7 Governance

BioExcel’s intention is to build a user-governed governance structure to ensure that strategic decisions are aligned with long-term needs of our users. Furthermore, the evaluation of impact generated through BioExcel’s work should be monitored by independent parties. This can be done by setting up advisory boards with representative stakeholders, who support the board of directors and provide input to the management of the CoE. The scheme below gives an overview over the main governing organs and how they are interconnected.



Scheme: Proposed governance structure of BioExcel.

**Board of directors:** The board shall consist of user representatives and representatives from BioExcel partner organisations. BioExcel will ask users from its community to join a virtual user council. Once a year this user council elects its representatives into the board of directors, which ideally should contain user representatives from academia and industry. The board will decide on the center’s strategy, and follow up with the progress made by the management. It is the responsibility of the board of directors to appoint members for the two advisory boards.

**Scientific advisory board:** Members of this scientific advisory board represent stakeholders from academia and industry with expertise on biomolecular research and software.

**Business advisory board:** Members of the business advisory board should contribute expertise in sustainability, innovation management and industry standards.

**Management:** The management team is responsible to oversee and coordinate the activities of the BioExcel partners, to ensure that the work aligns with the strategic goals, and to act as contact point to users and other stakeholders, to liaise with industry, and to attract funding.

## Appendix

### I. List of ISVs

The list contains companies that develop software for biomolecular research. Some of these companies also offer consultancy services around their software offerings.

Company name	Product or Service	HQ/Offices
Accellera	Molecular dynamics	Spain
Agile Molecule	Abalone modelling	Sweden
Applied Maths	Bioinformatics, biotechnology	Belgium
Astridbio	Data analysis, genome analysis	Luxemburg, Hungary
Bioblender	3D visualisation	Italy
BIOMOS	GROMOS dynamic simulations	Switzerland
BioPharmics	Computational drug design	-
Bio-Product	Protein engineering	Holland
BioSolveIT	FlexS/FlexX molecular modelling	Germany
Biovia	Modelling and simulation	UK
Cambridgesoft	ChemDraw, Chem3D, ChemFinder	USA
Certara	Small compound design software	USA
ChemAxon	Cheminformatics /Bioinformatics	Hungary
Chemical Computing Group	Molecular Operating Environment	Canada, Germany, UK
Chemotargets	CT-link for drug discovery	Spain
Cresset	Molecular design	UK
Dotmatics	Data visualisation	UK, Germany
Fujitsu	SCIGRESS technical computing	Japan, Poland
Gaussian	Computational chemistry	USA
GeneStack	Bioinformatics	UK
Hypercube	Hyperchem - molecular modelling	USA
IDBS	IT solutions for biomolecular res.	USA, UK
Instem	Life sciences IT solutions	USA, UK
Inte::Ligand	Ligand design and modelling	Austria
Lead Molecular Design	Drug design	Spain
Materials Design	MedeA modelling suite	Austria, Sweden
Mind The Byte	SAAS for drug design	Spain
Molecular Discovery	Drug design software	UK
Molsoft	Structure prediction, bioinformatics	USA
Novamechanics	Drug discovery modelling	Cyprus
OpenEye	Drug design software	US, Germany, France
Optibrium	Small molecule design	UK
Quattro Research	Data productivity workflows	Germany
Samson Connect	Computational nanoscience	France
Schrödinger	MD & QM software	USA, UK
Tremolo-X	Molecular dynamics	Germany
Wavefunction	Spartan – molecular modelling	USA
Yasara	WHAT IF molecular modelling	Austria, Netherlands

### II. National funding opportunities

Most European countries have governmental institutions and foundations that support academic research as well as industry-academic collaboration for innovation and valorisation projects. The list below contains examples of national funding



opportunities that BioExcel could access for single standing projects. The list is by no means exhaustive.

### **Netherlands**

1. The Netherlands Organisation for Scientific Research (NWO, [www.nwo.nl](http://www.nwo.nl)) funds top researchers and steers the course of Dutch science by means of research programmes, and by managing the national knowledge infrastructure.
2. The eScience Center, a joint initiative of the NOW and the Dutch organisation for ICT in education and research (SURF) is a national hub for the development and application of domain overarching software and methods for the scientific community ([www.esciencecenter.nl](http://www.esciencecenter.nl)).

### **UK**

1. Academic research funding is made available through the Research Councils UK ([www.rcuk.ac.uk](http://www.rcuk.ac.uk)). All councils have grants that can be applied for on a case-by-case base in particular themes. Councils potentially applicable to BioExcel's UK partners are:
  - a) Biotechnology and Biological Sciences Research Council ([www.bbsrc.ac.uk](http://www.bbsrc.ac.uk))
  - b) Engineering and Physical Sciences Research Council ([www.epsrc.ac.uk](http://www.epsrc.ac.uk))
  - c) Medical Research Council ([www.mrc.ac.uk](http://www.mrc.ac.uk))
  - d) Science and Technology Facilities Council ([www.stfc.ac.uk](http://www.stfc.ac.uk))
2. Several trusts distribute grants to individuals and research teams.
  - a) The Wellcome Trust ([wellcome.ac.uk](http://wellcome.ac.uk))
  - b) The Leverhulme Trust ([www.leverhulme.ac.uk/](http://www.leverhulme.ac.uk/))
  - c) The Carnegie Trust ([www.carnegie-trust.org](http://www.carnegie-trust.org)).
3. Governmental funding for companies and partner organisations is provided through Innovate UK ([www.gov.uk/government/organisations/innovate-uk](http://www.gov.uk/government/organisations/innovate-uk)). Consortia are required to be UK industry led, but in partnership with UK academic institutions.

### **Spain**

1. The Spanish Ministry of Economy, Industry and Competitiveness provides funding for postdoctoral research ([www.mineco.gob.es](http://www.mineco.gob.es))
2. The agency for Management of University and Research Grants (AGAUR) of the government of Catalonia provides scholarships and grants for academic research and knowledge transfer.
3. The Everis foundation ([www.fundacioneveris.com](http://www.fundacioneveris.com)) and the “la Caixa” foundation ([www.caixaimpulse.com](http://www.caixaimpulse.com)) both provide funding for valorisation and entrepreneurship.

### **Sweden**

1. The Knut & Alice Wallenberg foundation ([www.wallenberg.com](http://www.wallenberg.com)) promote scientific research with impact on the Swedish society.
2. The Wenner-Gren foundations ([www.swgc.org](http://www.swgc.org)) support international scientific exchange.

3. The innovation agency VINNOVA financially supports Swedish industry-academia consortia for valorisation and innovation projects ([www.vinnova.se](http://www.vinnova.se))
4. The agency for economic and regional growth, Tillväxtverket, distributes EU structural funds to support regional growth, which includes project funding for SME-academic collaboration.

**Germany**

1. The German Research Society (DFG) provides funding for academic research.
2. The German Federal Environmental Foundation of ([www.dbu.de](http://www.dbu.de)) provides funding for environmental research and innovation projects.
3. The Volkswagen Foundation ([www.volkswagenstiftung.de](http://www.volkswagenstiftung.de)) supports international collaboration in academic research.