INFRAEDI-02-2018

D5.3 - Business Model Design Report

WP5: Governance and Business Sustainability

Document Information

Deliverable Number	D5.3
Deliverable Name	Business Model Design Report
Due Date	2020-01-31 (PM13)
Deliverable Lead	AL
Authors	Rossen Apostolov (KTH) Angele Giuliano (AL) Ian Harrow (IHC) Richard Norman (NC) Steve Robertshaw (AL) Yvonne Westermaier (NBD)
Keywords	Sustainability, Market Modelling, User Needs, Business Offer, Infrastructure, Finance
WP	WP5
Nature	Report
Dissemination Level	Public
Final Version Date	2020-01-29
Reviewed by	Alexandre Bonvin (UU) Erik Lindahl (KTH) Xavier Salazar (BSC) Mark Sawyer (EPCC) Yvonne Westermaier (NBD)
MGT Board Approval	2020-01-25

Document History

Partner	Date	Comments	Version
AL	2019-12-03	First draft	0.1
KTH	2019-12-06	Executive summary and other edits	0.2
AL	2019-12-23	Revisions after internal review	0.3
AL	2019-12-28	Revisions after PMB review	0.4
KTH	2019-12-30	Additional edits	0.5
EPCC	2020-01-06	Additional edits	0.6
EPCC	2020-01-07	Additional edits	0.7
EPCC	2020-01-13	Additional edits	0.8
AL	2020-01-19	Additional edits	0.8a
EPCC	2020-01-19	Cleaned up version (no change tracking)	0.9
AL	2020-01-21	Merge of v0.8a and 0.9 with track changes still on	0.9a
KTH	2020-01-22	Minor edits and clean-up	1.0
AL	2020-01-22	Annex A replaced	1.1
AL	2020-01-25	Minor edits after MGT approval	1.2
KTH	2020-01-29	Finalized	1.3

Abbreviations

The following abbreviations are used in this document for the partners in the BioExcel project:

Full name	Abbreviation	Country
KTH Royal Institute of Technology	KTH	Sweden
Universiteit Utrecht	UU	Netherlands
Fundacio Institut de Recerca Biomedica	IRB	Spain
Forschungszentrum Juelich	JUELICH	Germany
The University of Edinburgh	UEDIN	UK
The University of Manchester	UNIMAN	UK
University of Jyväskylä	JYU	Finland
European Molecular Biology Laboratory	EMBL-EBI	Germany
Max Planck Gesellschaft	MPG	Germany
AcrossLimits	AL	Malta
Barcelona Supercomputing Center	BSC	Spain
Ian Harrow Consulting	IHC	UK
Norman Consulting	NC	Norway
Nostrum Biodiscovery	NBD	Spain

Table 1: Organisation Abbreviations

Executive Summary

One main objective of BioExcel as a publicly fully funded project is to exploit opportunities for commercialization of products and services stemming from its activities, as well as identifying a range of funding schemes for long-term sustainable operations. This deliverable presents the work undertaken to date with respect to inspection of a viable business model based on the experiences from research conducted in the first phase of BioExcel. To avoid repetition, we have not included results and conclusions already presented in previous deliverables (see footnote 1 on page 6).

Using the widely adopted Business Model Canvas tool, we analyse the value chain of infrastructure, offerings, customers and financing. Workflows (WF) solutions (generic and custom ones), which package core applications with auxiliary tools for molecular modelling, have been identified as products that are highly needed by the wider communities. They improve the productivity of researchers, allow for efficient utilization of compute resources and lower the time to solution of complex scientific problems. The value of the WF products is considerably increased by providing alongside them a catalogue of complementary training services which have already reached Technology Readiness Level (TRL) 7 to 9 through BioExcel efforts. Market analysis shows the pharmaceutical industry as a major potential customer, with a market size sufficiently big to support commercial offerings that will be sustainable. User surveys and focus group meetings with industry representatives showed concrete areas of interest such as virtual screening, current issues with existing software solutions (e.g. price, long-term sustainability considerations, trustworthiness of results when using closed-source tools, etc.) and highly desirable tools and features (e.g. automation, performance, expertise, etc.). We know, from the S&T partners in the consortium, that many of the large pharma companies use some of the software tools brought into BioExcel by these partners. However, it is important to note that in addition to this existing usage, we have seen further concrete interest from industry to access BioExcel expertise and other tools which are not purely part of the molecular modelling domain, particularly use of the Common Workflow Language (CWL) in commercial environments.

Consortium partners are in position to provide the necessary resources in terms of expertise and tools for delivery of the offerings during the project lifetime. Moreover, an extensive network of partner contacts as well as BioExcel agreements with third parties bring additional leverage to our activities. Effective delivery of effort towards commercial activities is however subject to various constraints from the consortium partner institutions. It has been challenging to identify a suitable governance model for the legal body that accommodates the existing complex requirements. We consider Social Enterprise as the best choice of legal entity, formed as a virtual organization, and are taking the necessary steps towards its establishment.

The work presented here will feed into the business planning activity which is the next exercise to be undertaken towards the planned deployment of a commercial operational arm of BioExcel CoE.

Contents

1 Introduction

Initial sustainability studies were undertaken in the first funding phase of the BioExcel project[1] to determine whether or not the CoE might be able to achieve and maintain marketplace operations. There, the focus was on what the consortium might be able to deliver to the market and to try to understand where income streams might be discovered. It was well understood that cutting-edge research on extreme-scale scientific computing will need to be supported to a large extent by public funding. However, there exist opportunities for commercialization of specific products and services, which can support the operations of a commercial entity.

In the second funding phase, we continue building on the previous achievements. After having established quality mark and branding guidelines (BioExcel-2 deliverable D5.1 - Branding and Quality Mark Design Inception Report), we made a full inventory of business exploitable assets (BioExcel-2 deliverables D5.2 - Annotated and Justified Service Catalogue). Now, we turn our view out towards the market in order to understand user needs, to model market dynamics, and to assess the viability of any potential opportunity.

We have determined that even under the worst-case market modelling scenarios, there is an opportunity for BioExcel. Moreover, we have discovered unmet user needs that we can address in order to capitalise upon the opportunity. It is to be noted that these unmet user needs map very well onto the offerings we have determined as being most credible for BioExcel to offer as commercial services.

In identifying the market opportunity, we have determined that the most suitable initial business model for the commercial arm of BioExcel CoE is that of a *Biomolecular Modelling Community Support & Development Service.*

The form of legal entity we plan to establish in order to manage the commercial operations of the BioExcel CoE is a Social Enterprise, which is to be established in Sweden. The company will form virtually (see Section 6.2.1) to keep initial costs low and will grow organically. The company will operate for profit in the commercial marketplace. Any income will be used to maintain the operational capacity of the legal entity and any surplus will be reinvested in the scientific community that it supports. The precise mechanism for doing this is to be determined.

It should be noted that there is a living document which forms the basis for this deliverable. The living document will be continually updated as more information is acquired and incorporated. Future deliverables will be informed by and built upon by this living document.

---

[1] D5.1 – Initial Business Plan [doi:10.5281/zenodo.263932], D5.2 – Revised Business Plan [doi:10.5281/zenodo.574634], D5.3 – Updated business plan [https://zenodo.org/record/3564172], D5.4 – Governance and Business Plan [https://zenodo.org/record/3564177]

2   Business Model Canvas

The Business Model Canvas (BMC) first described by Osterwalder and Pigneur in 2010[2] is a visual template which has been used by numerous entrepreneurs and start-up companies to develop an idea into an operating business. By focusing on and developing a set of interconnected areas required for business success the BMC provides guidelines for business prototyping. The BMC features nine business blocks: Value Propositions, Customer Relationships, Channels, Customer Segments, Key Activities, Key Partners, Key Resources, Cost Structure, and Revenue Streams. These have been consolidated into four focus areas (Figure 1) which will be used to guide the design and development of the BioExcel Business Model as well as provide the structure for the rest of this deliverable.
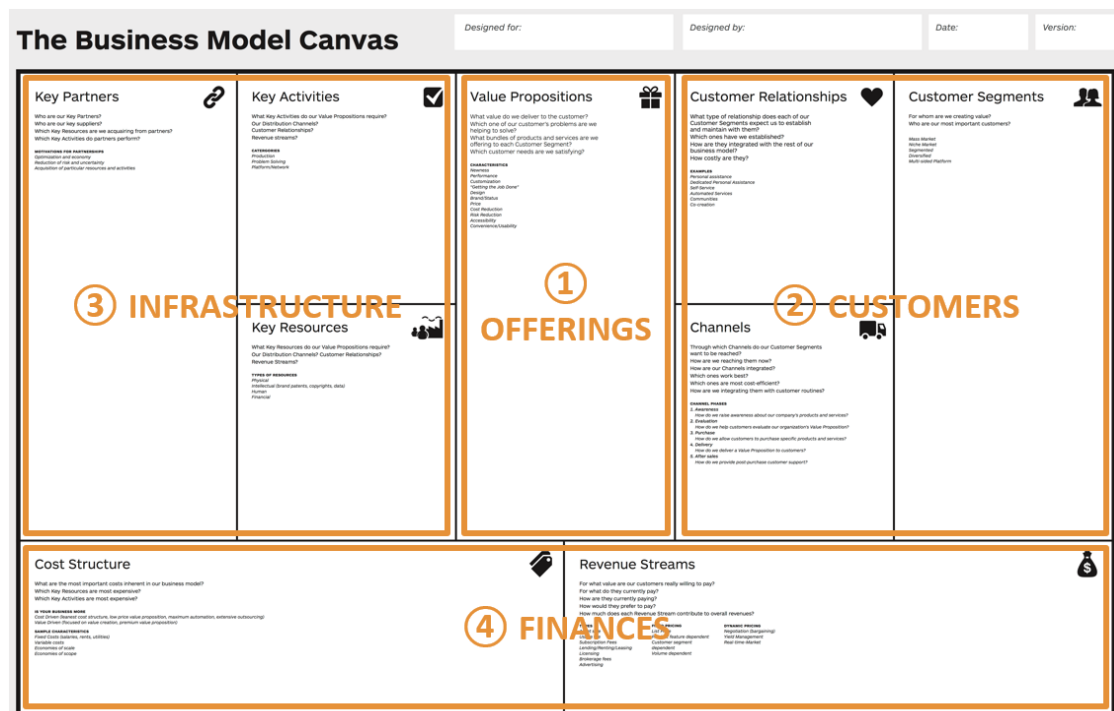

Figure 1. The original BMC featuring the nine business building blocks with the four consolidated focus areas highlighted: ① Offerings, ② Customers, ③ Infrastructure and ④ Finances.

Each focus area is described in more detail below and our work in each of these areas is described in subsequent sections.

Offerings is concerned with defining the value proposition and demand for BioExcel. This area of focus will involve consolidating the BioExcel value proposition elements versus the competition and understanding the importance of these on the 'job to be done' from a customer's perspective. Some of the questions we will be looking to answer include: which challenges are being faced by users and which solutions are being used currently to address these?

Customers is concerned with defining the volume and value of the market for BioExcel's value proposition. This area of focus will involve consolidating known industry sectors relevant to BioExcel, e.g. pharmaceutical, and explore other

---

[2] Osterwalder, A., Pigneur, Y., In Clark, T., & Smith, A. (2010). *Business model generation: A handbook for visionaries, game changers, and challengers.*

sectors as potential customers. Some of the questions we will be looking to answer include: which industries and sectors use molecular modelling and how many organisations and users does each have?

Infrastructure is concerned with defining the resources and support network required to establish and operate a successful business. This focus area will involve consolidating the partners (personnel and associated skills), resources (e.g. software and hardware), and the key activities (e.g. customer relationship management) required for the delivery of BioExcel's value proposition. In addition, this area will consider which additional partnerships, deals and alliances need to be established and formalised to enable the value proposition. Specific to this is the creation of a Legal Entity to support commercial operations.

Finances is concerned with defining the income and cost streams to prioritise the BioExcel cost structure. This focus area will consolidate all potential income and cost streams to create a financial model based on estimated figures for customer demand, pricing, probability of success of securing funding, etc. A good initial understanding of Customers and Infrastructure is required before attempting to define the financial model. The model will be refined over time and used to establish the business cost structure and price point for its offerings to ensure viability and sustainability.

3   Offerings

The basic question underlying our offerings is how the molecular modelling capabilities of BioExcel tools can be transformed into solutions for industry users. Therefore, our offerings need to be client centred. An initial list of business-exploitable assets as of September 2019 is provided in Deliverable D5.2[3] - Annotated and Justified Service Catalogue. We define a business-exploitable asset as any value proposition of a sellable or marketable service or product by the company to be founded. We define three main categories of assets: (i) workflows, (ii) training, and (iii) solutions (packages), and provide details for each of them. Molecular modelling solutions combine the workflows and training assets as solutions or packages that can be adapted to the needs of the client. The assets will be further developed throughout the BioExcel project, updated at least every six months, and prioritized in later deliverables (D5.4 Preliminary Business Plan available at PM18 and D5.5 Final Business Plan at PM30).

The workflow assets are listed in Table 1 of D5.2[3] and include use case-based workflows (antibody design, high-throughput modelling of human interactome, rational drug design, and electronic interaction phenomena), subtasks of these workflows, customized workflows (based on customer demands), and pipelining of basic building blocks. Means to improve the user experience are for example a KNIME Analytics Platform GUI, a web server GUI for fixed workflows, or HPC as a service.

---

[3] D5.2 has been delivered to the EC but is yet to be approved at the time of writing of this document. This footnote will be updated to replace this current text with a DOI for D5.2 when it is approved.

The training assets can have three different modalities: on-line, face-to-face or tailored to specific needs of a client. The modules or courses on BioExcel tools are listed in Table 2 of D5.2[3], and are updated at each major release of a tool. Recent training examples include trainings on BioExcel applications to industry users (HADDOCK training at UCB and in collaboration with Janssen on PMX), the BioExcel Summer School, and the PRACE/BioExcel Seasonal School.

Rough estimations of the TRL were used to assess how close we are to offerings that can be commercialized.

On the workflow side, the most advanced one is on virtual screening and is at an estimated TRL of 5 to 6 (validated or demonstrated in a relevant environment). However, the majority of workflow assets are at a TRL of about 3 (proof of concept stage). They cover the following themes: antibody or enzyme engineering, study of the human interactome, lead optimization of small molecules or peptides, discovery and optimization of protein-protein interaction inhibitors, design of "function-recovery" drugs for tumour suppressors, development of a scoring function tailored to a particular protein target, optimization of fluorescent proteins, and study of proton dynamics. Importantly, a large gap will have to be closed to reach workflows that can be commercialized.

On the training side, most of the assets are at TRLs of 7 to 9 (prototype or actual training proven in operational environment). An exception is CP2K with an estimated TRL of 2 (technology concept formulated; newly incorporated into the project in 2019). The commercial pricing range for half-day or one-day training will need to be defined: e.g. a fixed price for standard modules and additional payment for customized solutions. We have provided initial price estimates previously in the BioExcel-1 deliverable "D5.4 - Governance and Business Plan"[4].

In the end, we aim at tailoring molecular modelling solutions for industry users based on our offerings. The possible solutions will need to be centred on client needs and will span a whole spectrum of workflow and training combinations: Installation or assistance with Exascale computing, training of future users, or all-in-one solutions (installation, provision of customized workflows, training, and long-term consultancy). We will undertake measures to prevent security, confidentiality and privacy issues associated with running Software as a Service (SaaS) solutions. Where confidentiality is not an issue a range of free services already exist on the BioExcel website (e.g. training, Knowledge Resource Centre with tutorials, webinars, support forums, and cloud portal).

4   Customers

BioExcel CoE users represent a market opportunity for the commercial arm of BioExcel. In this section we consider both the user as a potential customer, by gaining an understanding of their needs; and the market, in terms of gaining an understanding of its size and dynamics. Jointly, these two exercises provide us with sufficient awareness of the opportunity to determine if it is likely to support an independently sustainable business operation.

---

[4] https://zenodo.org/record/3564177

Our current view is that the global molecular modelling market will be the main market for the BioExcel value proposition, with the pharmaceutical industry seen as providing the largest opportunities for commercialisation. This market is our focus in this deliverable. We are aware that there are wider opportunities for BioExcel but these will require new markets to be accessed and explored. These opportunities will be considered in future work.

Below, we present an initial analysis of market conditions and future projections and a preliminary assessment of user needs that have emerged from a number of customer-facing activities: A preliminary User Needs Survey, several Focus Group Meetings, the feedback collected from face-to-face interviews conducted during a number of site visits and the industry outcomes from a Free Energy Workshop the project organised and delivered. The questions used in the User Needs Survey are attached in Annex A to this deliverable. For confidentiality considerations, a graphical summary of the collated answers to the User Needs Survey will be provided for review as a complimentary document in addition to this deliverable.

4.1   Market Analysis

An internal market analysis has been undertaken. As this kind of analysis is based on an open engagement exercise rather than a scientific process, we report it here is a manner that is more narrative than is common in project deliverables. However, this narrative approach captures all stages of the process and accurately reveals the findings.

The initial plan for us to gain an understanding of the global molecular modelling market and the opportunities it represents was to purchase a market forecast from one of the many specialised market research companies that exist. It is usually the case that such companies will provide outdated or partially redacted reports to demonstrate their credibility to potential clients. To our surprise, we were not able to obtain samples that were of any use in assessing credibility, e.g. Molecular Modelling Market Research Report-Forecast till 2023 by Market Research Future[5]. We attributed this to the impression that the existing market forecasts are based on naïve assumptions about the Biomolecular Modelling market. This is based on the observation that on many websites, where market forecasts are promoted, the companies selling the reports cite sources that do not fit well with our understanding of the companies operating in the molecular modelling market. It is not a case that companies have been missed, but rather that companies that operate in adjacent markets are included as sources. This caused us to doubt the methodologies used and we decided to continue our search for alternative sources. We found very little of any concrete value or leading to any reasonable level of confidence. We found only three data points from three different sources: a historical global market value from 2017, a forecast market value for 2023, and a compound growth rate from 2018. Notably we observed that despite the low confidence we have in the market data, all sources we found agreed that the biomolecular modelling market is growing rapidly.

[5] https://www.marketresearchfuture.com/reports/molecular-modelling-market-4330

We found ourselves in the position of having a complete lack of confidence in the relevance of any forecast we might purchase. Rather than rely on what we considered to be poor or incomplete forecasts, we decided that it would be better to generate our own forecast based on our domain expertise.

Using the three data points from three different years mentioned previously, a simple regression line was calculated to give a simple model of the market. This was used to derive further values to divide the market into segments relevant to BioExcel. Firstly, we modelled that part of the global market occupied by companies operating in the software and services sector. Again, these data were obtained by reviewing market information openly available on the web. Subsequently, we took a best- and worst-case view of what part of that market was not dominated by (relatively) large companies in biomolecular modelling market. This information was taken from Owler[6], a crowd-sourced company valuation website, and minimum and maximum values were taken for each relevant company. These values were totalled and used to generate two more global segments. These segments represent the maximum value of global opportunity for BioExcel. See Figure 2 below.
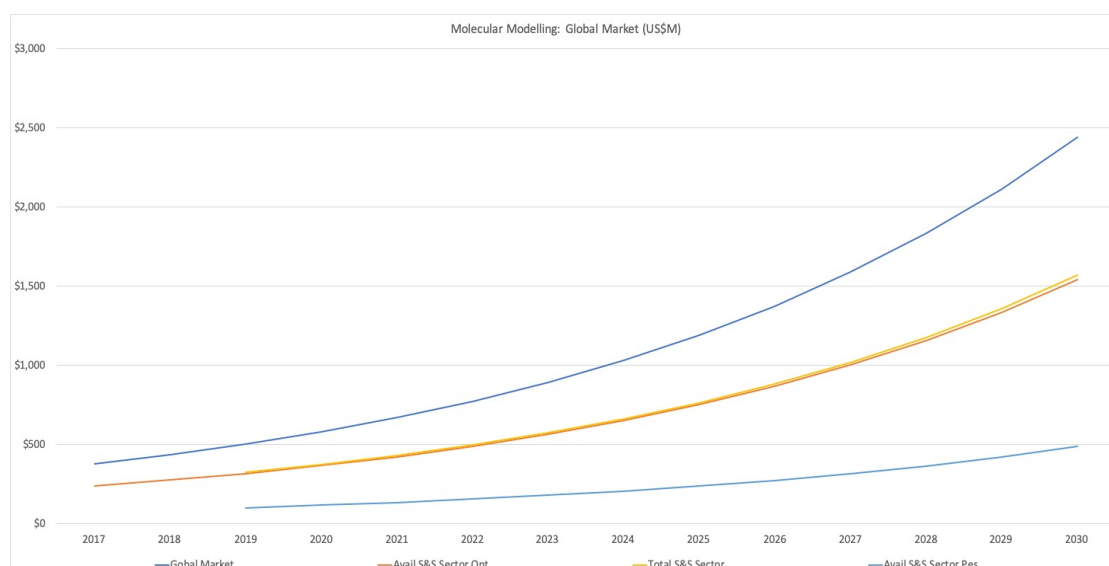

Figure 2: BioExcel Global Molecular Modelling Market Opportunity. Abbreviations used: Avail (Available), S&S (Software and Services), Opt (Optimistic), Pes (Pessimistic).

However, for several factors (market pressure, investment, staffing, etc.), BioExcel is not reasonably going to reach that level of market presence. We, therefore, continued to decompose the global model, firstly into a regional model using EU data from Eurostat[7] which will be the first focus for the BioExcel CoE. From this regional model we generated sub-models that represented our Serviceable Available Market (SAM) and Serviceable Obtainable Market (SOM) in both best- and worst-case conditions. Although the worst case is not encouraging for rapid growth, it is not a negative value and it indicates there is a market for the BioExcel

---

[6] https://corp.owler.com/

[7] https://ec.europa.eu/eurostat/web/main/home

CoE to tap into and to grow organically whilst supporting itself. See Figure 3 below.
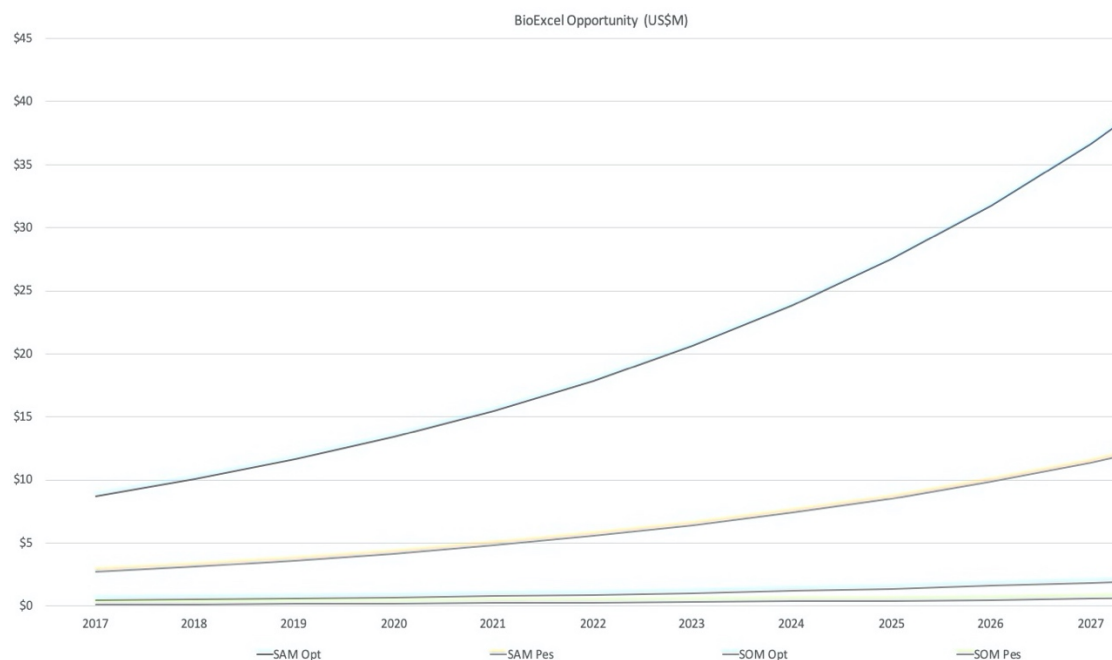

Figure 3. BioExcel Best- and Worst-Case Opportunities after Regionalisation.

It must be re-emphasised that the modelling carried out to date is preliminary and the forecast values really are based on the most pessimistic of projections. The forecasts consider our main competitors but no actual analysis of their offers relative to our own has yet been carried out; this work will be undertaken in the forthcoming project period. However, the models discussed here are real and will be developed and refined in order to work towards improving the model, as more reliable data become available. Despite these caveats, we are pleased that there is a market opportunity for BioExcel CoE. The opportunity expresses the limits of what is possible for growth, it is not an indication of income or turnover. Despite the likelihood that the space we have discovered will grow over time as the market itself grows, the commercial arm of BioExcel will grow organically[8] within it and, once established, will explore various means to increase the space to allow further forms of growth to take place.

Data used in market modelling[9]
Global Market Value 2017 $377M[10]

---

[8] Growth that takes place as a result of re-investing all or part of profit, as opposed to borrowing money or attracting venture capital.

[9] Please note that market values are measured in US dollars, not Euro: US dollars is the de facto global business currency. Meaningful comparison is not possible when converting between currencies because of fluctuating exchange rates and the market view of risk differentials.

[10]   https://www.marketwatch.com/press-release/molecular-modelling-market-2019-global-industry-demand-recent-trends-size-and-share-estimation-by-2023-with-top-players---marketreportsworldcom-2019-05-02

Global Market Forecast Value 2023	$893.1M[11]
CAGR[12] Discovered Values	15.16%[11] 15.25%[13] 15.46%[10]
CAGR Derived from regression	15.44%
S&S Market Share	63.10%[14]
Taken S&S Sector - Optimistic[15] (2019)	$180M[16]
Taken S&S Sector - Pessimistic[17] (2019)	$402.5M[16]
EU Share of global Market	27%[18]
SAM*	10%
SOM**	5%

*adjusted for BioExcel offerings, i.e. workflows, training and packages
**based on fraction which can be captured by BioExcel from competitors
NB: Both the SAM and SOM values here are arbitrary. We will refine these worst-case values as we gather more data.

4.2   Initial Industrial User Survey Results Analysis

The findings discussed below in Section 4.2 have proven to be contentious. Readers should pay particular attention to footnote 21 on page 15, to the Note Box at the end of Sub-section 4.2.3 on page 17 and to the constraints described in Sub-section 4.2.7 on page 20.

The results, which are analysed here were collected from a survey of industrial users of biomolecular research tools. In total, there were eleven questions, but the number of questions actually answered depended on the route taken through the complete set; routing was determined by the answer given in one question. The questions, which are found in Annex A, themselves were developed iteratively over a period of several weeks, with feedback being received from domain experts within the consortium.

A decision was taken to allow respondents to complete the survey anonymously. We reasoned that respondents were more likely to answer the survey honestly if they were able to remain anonymous. However, this decision may have led to unintended consequences, which we address later.

---

[11]   https://www.marketwatch.com/press-release/molecular-modelling-market-size-industry-growth-share-opportunities-emerging-technologies-competitive-landscape-future-plans-and-global-trends-by-forecast-2023-2019-11-04

[12] Compound Annual Growth Rate

[13]   https://galusaustralis.com/2019/11/81791/molecular-modelling-market-growing-by-cagr-of-15-25-with-certara-genedata-schrodinger-indigo-instruments-advanced-chemistry-development-inc-thermo-fisher-scientific-dassault-systemes-physiomi/

[14]   https://www.marketwatch.com/press-release/molecular-modelling-market-and-top-key-players-are-imulations-plus-inc-chemical-computing-group-certara-lp-openeye-scientific-software-2019-08-20

[15] After this sum is removed from the global value, the remainder is available for other services

[16] https://corp.owler.com/

[17] After this sum is removed from the global value, this larger remainder is available for other services

[18] https://ec.europa.eu/eurostat/web/main/home

There were 39 responses to the user survey, which were collected from two campaigns. The first campaign was initiated by direct messages to individuals already connected to consortium members, the second campaign was initiated through open social media invitations. The survey remained open for two months.

In this preliminary analysis of the user feedback collected during the user survey, we focus exclusively on segmentation by company size. Segmentation by market sector is largely meaningless due to the overwhelming representation of Pharmaceutical companies in the respondent list. In order to understand whether differences in user needs arise due to factors associated with market sector, we will make strenuous efforts to gather data from a wider sample of users in future iterations of the user survey.

4.2.1   Company Characteristics (Q1 - Q4)

Only respondents from Micro, Small and Large companies[19] replied. No responses were collected from representatives of Medium-sized companies.

We collected responses from representatives of ten Micro companies, five Small companies and eighteen Large companies. Of these, twenty were multinationals (one Micro, two Small and seventeen Large). By far, the majority of companies employed their own modelling teams (Micro - 10 out of 11, Small - 4 out of 6, Large - 21 out of 22).

These companies operated mainly in the pure Pharmaceuticals sector, but the second largest cohort worked jointly in the Pharmaceutical and Biotechnology sectors. A small number of respondents represented companies working in other specialised areas. By company size these distributions held up quite well as follows:

Micro
Mainly pure Pharma, with three also working in Biotech and one additionally working in Speciality Chemicals, Petrochemicals and Cosmetics.

Small
Mainly Pharma/Biotech, with one working purely in Pharma and one additionally working also in Speciality Chemicals and Agricultural chemicals.

Large
Mainly pure Pharma, with four also working in Biotech (including one also working in Speciality Chemicals) and another focusing exclusively on Agricultural chemicals.

4.2.2   Challenges (Q5)

Overall, the challenges faced by respondents were reported to be relatively evenly distributed through the challenge set we offered, with Molecular Dynamics, Virtual Screening and precise binding energy prediction forming the top three.

---

[19]   https://ec.europa.eu/eurostat/web/structureal-business-statistics/structureal-business-statistics/sme

However, the distributions do differ between the Micro, Small and Large companies. For the Micro companies Virtual Screening and Molecular Dynamics significantly represent the focus of interest. For the Small companies, Virtual Screening and Off-Target Modelling represent the focus of interest. For Large companies, the focus is on precise binding energy prediction and Molecular Dynamics, with Virtual Screening and Molecular Determinants following closely behind. Of the challenges represented in our answer set, we missed enzyme engineering, homology modelling, and ADME[20].

4.2.3 Satisfaction (Q6)

In terms of overall satisfaction[21] with tools and services(defined here as how how well the tool in question met user needs), the median varied only one point between the commercially offered tools and services and those produced by consortium members (Very Satisfied vs. Satisfied); moreover, the range from which these median values were extracted were also very similar. In fact, the lowest satisfaction values are associated with a commercial tool[22]. So far, as "support services" are concerned, there seems to be an overall level of satisfaction with the various training and consultancy services listed.

However, things start to look a little different when the responses are clustered by company size.

Micro companies are similarly satisfied with the Open Source tools as they are with the commercial tools; however, they are less satisfied than the overall picture with the "support services", some citing cost as a barrier, e.g. for commercial tool training, with on-site training and online training receiving the lowest satisfaction.

Small companies are also similarly satisfied with GROMACS as they are with the commercial tools; however, they cite slightly less satisfaction with both on-line and locally installed versions of HADDOCK and are dissatisfied with PMX. They are not satisfied with the "support services" they receive in general, being only very satisfied with their online training. Satisfaction with the other "support services" is rated at Satisfied or lower level. This satisfaction with support services is true across the entire range of tools in use, with deepest dissatisfaction noted against the commercial tools: we assume because of the costs associated with accessing them. This reveals a potential commercial opportunity for BioExcel commercial

---

[20] Absorption, Distribution, Metabolism and Excretion, the acronym ADME describes the main processes involved in pharmacokinetics.,

[21] Satisfaction is a common measure employed in user surveying, it does not measure the quality of the product, merely the user perception of it set within their overall experience and individual needs at that moment. For example, if someone has purchased a washing machine that is too small for their needs, the machine may itself be perfectly matched to all of its design criteria but from the user perspective, they would like it to be larger. This user desire is what we are trying to collect in order to plan what BioExcel takes to market and to feed back into the development teams to help us determine what divergence there may be between what users want and the research goals of the development teams. It is the job of the BioExcel legal entity to work out how to fill this gap, leaving the core developers to pursue their research goals. **In this case, we chose to measure user satisfaction over a single issue: that of how well the tool in question met user needs. No other aspect of the functional or non-functional performance of the tool should be inferred.**

[22] We deliberately withhold that commercially sensitive information here.

activities; through the value proposition associated with high quality support services wrapped around free to use tools. This could be a big hot with small companies.

Large companies are also similarly satisfied with GROMACS as they are with the commercial tools; however, they cite dissatisfaction with both versions of HADDOCK and especially PMX. They are Very Satisfied with the Workflow Consultancy and On-site Training "support services" but are only Satisfied with Workflow Training and Online Training "support services".

This level of analysis reveals that varying opportunities lay in the areas of dissatisfaction in each size of company. Micro companies seek affordable "support services". Small companies seek improvements to PMX and need improved Workflow Training and Online Training. Large companies seek improvements to HADDOCK and PMX and need improved Workflow Training and Online Training. One respondent ran local PMX and HADDOCK installations due to their company policy not allowing access to external services, a situation that reflects well on the deployment flexibility of BioExcel tools.

Respondents noted the following tools in addition to those already listed in the survey question:

- ICM and VMD which were not assessed.
- Open Eye Pipeline Pilot which was not assessed.
- AMBER, the respondent declaring it assessed them as a 4 out of 5 (very satisfied)
- Cresset Spark and Forge, the respondent declaring them assessed them both as 4 out of 5 (very satisfied).
- Acellera HTMD/MSM, the respondent declaring it assessed them as 3 out of 5 (satisfied)for MD, but stated that it was expensive
- Stardrop, the respondent declaring it assessed it as 3 out of 5 (satisfied) for SAR
- Seven anonymous tools/services, four evaluated at 3 out of 5 (satisfied), three were evaluated at 4 out of 5 (very satisfied). These tools are specifically not BioExcel tools.
- Seven other anonymous and un-evaluated tools/services were mentioned.

NOTE

During consultation, some scientific partners detected anomalies associated with some of the answers to the industrial user survey, especially those associated with question 6. Question 6 asked those respondents who had used a tool to measure the level of their satisfaction when using it.

Suspicions have arisen because the number of respondents assessing tools do not match what the tool owners know about the distribution of these tools. For HADDOCK, the concern is associated with the number of respondents assessing it is larger than the number of current industrial licenses issued. For PMX, the issue is associated with the proximity of the first viable release being issued and the period of the survey. Clearly, these issues raise concerns regarding the trustworthiness of the survey results.

Given this deeper knowledge of the current number of valid industrial licenses and access to software tools, we cannot distinguish between the following possibilities:
- *There are many individuals using and assessing each installation.*
- *Some industrial users are using the software with an academic license.*
- *The question was not understood correctly, leading to misinformation being entered.*
- *Deliberate misinformation has been entered into the survey.*

Our decision to allow anonymous survey responses contributed to our inability to determine what lay behind this anomaly.

We will work harder to ensure this situation does not arise in future surveys.

4.2.4   Features Required (Q7)

Overall, it is the technical aspects that most respondents seek to address, with Faster, Quality and Scale of Predictions being popular areas that need to be addressed, judging by the interest in HPC solutions, it seems that this type of computing is seen as a possible solution to these kinds of problems. However, also scoring highly overall is the need for improved tutorials and documentation. The least sought-after feature is application consulting. This may mean that this requirement is already adequately covered by the application providers or the design languages of the interfaces are sufficiently transparent to avoid the need for specialist support. However, the need for better user training and the need for improved user experience militates against this view; this tension has no obvious resolution. Better technical support and lower cost are only of moderate interest overall. Interoperability standards are not recognised as being of special interest, nor is the need to avoid provider dependency.

However, this picture changes significantly when looking at the data through the lens of company size.

Micro companies are also interested in improved technical performance, but they are much more interested in better tutorials and documentation as well as user

training and reducing costs than the overall picture suggests. HPC solutions are also of interest, again probably in association with the technical improvements. They are least interested in avoiding provider dependencies. NB: Those respondents seeking better tutorials and documentation usually also sought out better technical support.

Small companies are also interested in improved technical performance but mainly in the quality area and are interested in HPC solutions. They seek better tutorials and documentation and better technical support. There is a minor interest in application consulting, lowering costs, and reducing provider dependencies.

Large companies are overwhelmingly interested in improving the technical performance of their tools/services and have a large interest in HPC solutions. Despite this, there is also a strong interest in better tutorials and documentation and improving the user experience. Technical support and cost reduction are not a priority for Large companies, nor are interoperability standards, user training, and reducing provider dependencies. Application consulting/training is of little to no interest.

In addition to the features included in the question, the respondents listed the need for a service to build robust protocols / workflows to run FEP and other MD applications.

NB: It is clear from the way that many responded to Q6 that it was too complicated and we should seek to simplify the structure of questions in the forthcoming campaigns.

4.2.5   Willingness to Pay for Services (Q8 – Q10)

Overall, out of 39 responders, only two would not pay for services offered by BioExcel, one citing the belief that OS software and its service ecosystem should always be free of cost. Eight respondents declared that they would pay for our services. Twenty-two respondents declared that they would conditionally pay for our services.

The justifications cited by those willing to pay conditionally were:
- Price / Budget
    - Budget planning
    - Procurement approval
- Added value of tool / service
    - Capability improvement
    - Benefit delivered
- Selection depends on the outcome of a competitive evaluation with other tools
- Compatibility with other tools
- Business priorities and requirements
- Enterprise stability (purchase needs to be agreed across all groups)
- Training package included in price (clearly thinking of tool / software service)

- Licence model for infrequent use

There was no significant difference between respondents, based on company size, in relation to the way this question was answered.

NB: Large companies tended to be less interested in price than the Micro and Small companies.

The summary response data are contained, with collated answers, in a Confidential Annex to this document. The Confidential Annex is complimentary to this deliverable and can be accessed by those with the necessary authority.

4.2.6   Initial User Survey Findings

Of those responding to the survey, the typical industrial user profile is that of someone working in a dedicated molecular modelling group for a multinational, large pharmaceutical or biotech company.

The main challenges they face are virtual screening for lead discovery, dynamics of molecular complexes and binding energy prediction in lead optimization and off-target molecular modelling. Most of these challenges have some coverage in the BioExcel uses cases 1 to 3. Virtual Screening features in the top 3 challenges irrespective of company size. Free energy calculations seem to feature more prominently in large companies.

The BioExcel online PMX service could be improved. The complexity of local installation for HADDOCK and GROMACS is perceived as a barrier[23].

While overall workflow (of those listed in the interview questions) and training uptake is about 50-60% irrespective of company size and whether it is face-to-face or online, other discrepancies come to light when viewed from the perspective of company size. Generally, there is a mixed picture of satisfaction with all current workflows and training solutions. Those working in Large companies seem to be generally pleased with their tool support and training. However, those users in Micro and Small businesses express a clear need for affordable training and support. Such need represents an opportunity for BioExcel and fits well with our initial offer.

Competing solutions performed well, especially the market leaders, Maestro from Schrödinger and MOE from CCG, both of which registered high levels of satisfaction from this sample of respondents. BioExcel tools broadly registered similar levels of satisfaction with commercial tools, regardless of company size. It will be important to understand where improvements need to be made. Within the limitations of this preliminary survey, it seems that large companies reported an overall dissatisfaction with some BioExcel solutions; however, there are some concerns about the validity of this aspect of the survey input[24].

---

[23] For important additional context, please see the Note box at 4.2.3
[24] For important additional context, please see the Note box at 4.2.3

Large companies clearly prefer commercial tools to BioExcel tools; whereas, micro and small companies use BioExcel and competitor solutions fairly equally, reporting only a slightly higher satisfaction for competing solutions.

4.2.7   Constraints of the Survey

The survey was completed by 39 respondents. This relatively small sample size places constraints on the confidence we can attach to our conclusions. This issue will be addressed in future:

- by undertaking further surveys; from the trends we have observed during this analysis, we will derive the follow-up questions for future use.
    - During such surveys we will strive to ensure that users only respond to questions about tools they have actually used. As previously noted, there are suspicions that some of the results we have gathered in this preliminary survey have been skewed by users, for whatever reason, incorrectly answering questions.
- through conducting in-depth interviews with selected individuals and/or from additional site visits.

4.3   Focus group meetings

Focus group meetings are a means to get feedback on the relevance for industry users of the work done within the framework of BioExcel. Two focus group meetings have been organized so far, and both gave useful insights into how the development of BioExcel offerings can be driven by the end users.

4.3.1   Focus Group 1: London, November 2018

The first focus group meeting for industry users was held in London during November 2018. Three industry experts attended the meeting from the large pharmaceutical companies AstraZeneca, MedImmune, and UCB. Its purpose was understanding the needs and challenges for implementing molecular modelling in drug discovery and receiving feedback on a prototype of a building block-based workflow. The challenges highlighted fall into the following four categories: (i) expertise, (ii) software tools, (iii) time, (iv) security and intellectual property (IP).

The expertise gap (i) has to do with a lack of knowledge for evaluating the quality of a model and its underlying assumptions, and a need for more molecular modelers to support medicinal chemists. The industry experts bridge this gap currently with internal contractors, students, collaborations, and consultants/trainers mostly from three market leaders: Schrödinger[25], Chemical Computing Group[26] and 3DS Biovia[27]. Responsive on-site training such as that delivered by Schrödinger is highly appreciated. In contrast to most of these big molecular modelling software vendors, an advantage of BioExcel is that it also delivers a more holistic coverage of molecular modelling to educate the user. BioExcel can also deliver in depth training such as on HADDOCK which was seen as very useful and of higher quality than commercial offerings.

---

[25] https://www.schrodinger.com

[26] https://www.chemcomp.com

[27] https://www.3dsbiovia.com

Regarding software tools (ii), there is a need for user-friendly interfaces, using the KNIME workflow manager for automation beyond Python scripts. They also noted how Schrödinger's Maestro studio provides a very polished user interface which allows for a multi-seat licence at very high cost. Some of the industry experts prefer commercial tool providers over open source and academic providers for the quality of their very fast and responsive support.

With respect to time (iii), molecular dynamics is always more time-consuming than other molecular modelling tasks. External resources for faster calculations including cloud services are increasingly used.

Concerning security and IP (iv), external servers can only be used if security agreements and standards are certified. A complete due diligence would be needed to protect IP.

The participants suggested several improvements of the BioExcel solutions presented. The GUIs should be further developed, especially for non-expert users. Benchmarking and datasets for validation of the workflows are crucial. Security considerations should be included in the development. The BioExcel tools and workflows should allow for methods to be adapted and favour interoperability. Online tutorials for getting started or routine functionality would be useful.

The initial impressions on the prototype were positive: The building block approach coupled to KNIME seemed flexible and reusable, and outsourcing the web GUI maintenance could be attractive. The users liked the prospect of nodes for GROMACS and HADDOCK. However, fixed workflows will require responsive support like Schrödinger provides, local installation will need to be straightforward with few dependencies, and a utility library delivered through KNIME would be useful. The attendees would agree to pay for a BioExcel-supported solution once the tools are adopted and validated via benchmarks. If the pilot value is demonstrated, the budget for the solutions is likely to be forthcoming. This is already starting to happen with HADDOCK at UCB and PMX at Janssen.

Outcomes from this first focus group meeting include: establishment of a collaborative relationship with molecular modelers at AZ and MedImmune and the strengthening of an existing collaboration with UCB on PMX and HADDOCK. We also found potential interest in having access to high performance computing (HPC) infrastructure once the software tools have been adapted for molecular dynamics or docking at exascale using ensemble approaches.

4.3.2   Focus Group 2: Barcelona, June 2019
The second focus group meeting was held on the 26th of June 2019 at the Barcelona Supercomputing Center. It had a similar purpose but aimed at small to medium sized enterprise (SME) biotechnology and pharmaceutical companies.

Seven industry experts from SMEs, involved in molecular modelling and/or decision making attended the event. As a result of lively discussions, four key messages emerged: (i) Automation is challenging for stand-alone scientific

questions, (ii) the cost of a software does not necessarily justify its value/benefit, (iii) 'time to solution' is crucial; (iv) training is seen as very valuable.

Regarding automation (i), a crucial question is how BioExcel workflows can provide solutions to these scientific questions.

Two examples for cost of software (ii) were given: Schrödinger's FEP+ is very costly, and although good results can be obtained, measuring the Ki experimentally is often less expensive. However, Biovia's Pipeline Pilot is seen as an affordable solution for virtual screening in hit-to-lead and lead optimization campaigns.

With respect to the time to solution (iii), tools are trusted based on past and current use, while cross-validation is perceived a 'luxury'. Alternative solutions are considered based on competitive cost and demonstration of similar results.

The value of training (iv) is especially high for beginners in the molecular modelling field. Training is usually included in the price when purchasing new software. However, paying a fee for training and support is an option.

The industry users also gave several suggestions on how to improve the tools, workflows, and use cases. To lower the 'complexity' barrier, manuals for setting up, running and analysing the results are needed. The usability of the tools should be improved as part of the roadmap. To engage more users, there should be other cross-domain use cases e.g. on small molecules, cheminformatics, or crystallography.

4.3.3   Focus Group Conclusions

We learnt that future focus group meetings should provide a roadmap on what BioExcel will be for users in the long and short term: We aim to reach sustainability through formation of a social enterprise company which serves as a legal entity and business vehicle to deliver professional support services to industry users. These will need to include security standard protocols to ensure the protection of private data which has IP value for an industry customer. BioExcel also need to work out how focus group meeting attendees can access such services.

4.4   Site visits and interview feedback[28]

Site visits and individual interviews build on focus group meetings because they enable BioExcel to reach a deeper level of understanding user needs and challenges for a company. The large companies are likely to be multinational corporations operating across the globe with Computer Aided Drug Design (CADD) and/or Molecular Modelling (MM) groups located at different geographic sites. For example, UCB have CADD or MM groups in Belgium and the UK, Janssen, which is part of Johnson & Johnson have CADD/MM groups in Belgium and the US; whereas AZ and MedImmune have CADD/MM groups in Sweden and the UK. This complexity means that a visit to one site is likely to involve teleconference call-in

---

[28] Please note that the outcomes of these visits are reported in WP3 outputs. The visits are summarised here for completeness in coverage of efforts related to understanding BioExcel users.

from other sites during the visit, which helps BioExcel to build a coordinated relationship with each company. Besides this, individual BioExcel partners may meet with individuals from a company as part of an ongoing collaboration.

This year, there has been collaboration with UCB and Janssen on HADDOCK applications and PMX for free energy perturbation as follow-up to site visits during 2018. A third site visit took place at the AZ site in Gothenburg during October this year which focused on applications and new features of GROMACS which has resulted in a new collaboration.

Besides these site visits, such resultant collaborations have been facilitated by occasional teleconference or face to face meetings for informal interview feedback between BioExcel partners and individuals or groups. An example of such collaboration is the recent publication of a review article[29] which has been authored by BioExcel partners: Norman Consulting[30] and Utrecht University, together with industrial contacts, UCB, Boehringer Ingelheim, and NaturalAntibody.

A dynamic network of such collaborative relationships between BioExcel partners and pharmaceutical companies creates a strong position to build business relationships for delivery of professional services, by the BioExcel legal entity, when it is formed.

4.5   Free Energy Workshop and Industry Outcomes

Rapid and accurate calculation of the binding free energy between a small molecule or peptide and a protein target is of strong interest for drug or antibody design and engineering. Over 100 delegates from pharma industry travelled to Göttingen in Germany for the alchemical free energy calculations workshop[31] hosted in June this year by Bert de Groot's group, the core developers of PMX. Delegates included representatives from Bayer, UCB, Merck, Schrödinger, Janssen, and Boehringer-Ingelheim. This event is expected to be the start of a series of European workshops that will complement related future meetings arranged in the USA and help to transition free energy calculations into the mainstream.

An important outcome from this workshop for BioExcel and the delegates from industry is the recognition that community action is required to improve the interoperability between the different implementations of free energy calculations to avoid dominance by any single provider.

This could be tackled through formation of the Open Free Energy Consortium comprising of representatives from pharmaceutical industry and leading academic groups, including BioExcel partners. The nascent consortium has met two times since, mostly so that the common requirements could be agreed upon. This would then allow the academic and pharma partners of this consortium to work out a plan for implementation and options for funding in 2020. BioExcel sees

---

[29] DOI: 10.1093/bib/bbz095
[30] Through the Pistoia Alliance
[31] https://bioexcel.eu/bioexcel-alchemical-free-energy-workshop

this as an exciting opportunity to play a leading role as an active partner of this consortium which could result in publications and business services for professional support.

5 Organisational Infrastructure

In this section, we discuss the three main categories of infrastructure needed for provisioning of commercial offerings. They include key personnel and partners, activities in which they engage, and resources needed in support of those activities. Below we give a summary of each and discuss existing and envisioned constraints.

5.1 Partners

The current BioExcel consortium includes 14 partner institutions from 8 EU states. They bring expertise which covers the needs of both the current centre as well as its long-term operational commercial arm. Namely, scientific excellence and software for biomolecular modelling and simulations (KTH, UU, MPG, JYU, JUELICH, IRB, NBD); extreme-scale computing tools and infrastructure (KTH, MPG, JUELICH, BSC, UEDIN); automation and transversal tools and infrastructure (UMAN, BSC, EBI, IRB); training and support activities (EBI, UEDIN); and business development (AL, NBD, IHC, NC) (see **Error! Reference source not found.** for list of abbreviations). All partners have very extended networks with the world-wide scientific communities, the wider European HPC ecosystem including policy level working groups, as well as industrial alliances and networks.

The consortium is operating very well with high complementarity and leveraging of different skills. As part of the EU funding scheme we have dedicated effort for provisioning of services. However, a major constraint is the limited amount of this effort and the inelasticity of responding to higher demand. Hiring additional personnel on short notice for short- to mid-term assignments is challenging, particularly in academia. Moreover, all partners are subject to different regulations in their home countries which in some cases restricts the opportunity for provisioning of commercial services.

One way to address the gap is by establishing formal partnership agreements (contracts, possibly framework contracts) with host universities and institutes for time limited ad-hoc project-based collaborations. These collaborations will be funded at the institutes normal rates and within their limitations. The partnerships will be established with the legal body of BioExcel, which will serve as a facilitator in linking industry with academia. This facilitating role will provide faster and more scalable access to the expert services offered by the academic partners. Through a "train the trainer" model, the CoE will develop its own team of internal experts to offer the industrial user limited levels of service that do not impact on the academic partners. In that way, the time and effort to set up and execute a commercial request will be dramatically improved and effectively increase the uptake of academic expertise by industry and foster joint collaborative opportunities.

The first stage of BioExcel is to establish formal partnerships (via MoUs) with international collaborators (ELIXIR, INSTRUCT, MolSSI, HPC-Europa3, MuG VRE, Vi-SEEM, and Open PHACTS) and national leaders (Vrije Universiteit Amsterdam (VU) and Scuola Normale Superiore (SNS). Informal but strong collaborative partners include PRACE, the UK Software Sustainability Institute (SSI), the Pistoia Alliance, the US Food and Drug Administration, NIH Data Commons, EUDAT, EGI, Common Workflow Language community, Research Data Alliance, European Open Science Cloud, other CoEs (particularly FocusCoE, MAX, POP, and CompBioMed), CORBEL (biological and medical research infrastructures cluster project), and EOSC-Life. This collaborative network is expanding continuously as needs and opportunities arise.

5.2   Potential Commercial Activities

The main commercialisation products and services to be offered by us are custom workflow solutions, training/consultancy, as well as packaged solutions combining both. To ensure delivery of those, partners will have to engage in activities including but not limited to:

- outreach and close interactions with industry to understand their concrete problems and needs
- development of custom workflows, tools, containers, etc.
- providing technical support with deployment and execution of offered products, as well as fine-tuning and optimisation of the codes for given hardware resources
- providing scientific expertise through consulting sessions to ensure validity of the approaches and usage of the tools
- providing training through workshops, hands-on, site-visits, virtual platform, etc.
- providing access to compute infrastructure, either directly or through middleware interfaces (web, API, etc.)

5.3   Business Resources

There are five main resources needed for operations and delivery of products and services:

- scientific and HPC technology expertise
- full spectrum of software applications and tools
- compute infrastructure
- specialised knowledge
- training expertise

BioExcel possesses all the above but they are subject to a number of constraints. As mentioned in the "Partners" sections above, delivery of expertise on short notice is highly inelastic. To address this issue, we will initially utilise the available effort from within the project funding scheme to establish credible partnerships with potential customers. Building on those, we will aim to setup long-term joint projects and commitment from both parties (industrial users and academic providers), as well as attracting additional funding from third party investor bodies[32]. The centre has a wide spectrum of software but the intellectual property

---

[32] Such as: banks, investment banks, business angels, business dragons, venture capitalists, etc.

(IP) for the core tools stays with the respective partner who is developing them. Through agreements with the host institutions we need to ensure that those tools can be offered in a commercial setting, e.g. via appropriate licensing scheme(s). Access to high-end compute infrastructure with appropriate security standards is commonly desired by customers and some of the partner HPC centres offer secure access to their resources. Still, we need to establish the proper procedures and security protocols for each of the hosts since they are subject to different regulations by the host governing bodies.

5.3.1 Legal Entity

Through a legal entity, the BioExcel CoE will offer commercial services to customers (addressing their practical needs) with the aim to generate income that will enhance our support and development activities. support for academic users and enable us to improve the commercial offers. The process we are undertaking to achieve this goal is organic and will initially grow slowly. We ultimately plan that this work will lead to dedicated enhancements to the academic tools that make them more attractive to commercial users: enhancements such as training packages, GUIs, on-site technical support, etc. Clearly, training courses are easier to develop and this is where we will start commercialising the BioExcel services for market entry.

From our initial findings, we have determined that there is a continuing clear need for the services provided by the BioExcel CoE. The CoE has a willing team, willing initial investors[33], and interested subsequent investors.

For this reason, we have also taken the preliminary steps towards incorporating a legal entity to deliver such services.

Initiating the move towards company foundation has resulted in many obstacles being experienced. Notably, the type of legal entity was heavily constrained by the host organisations of many members of the consortium, many of which had limitations written into statute that prevented collaboration with certain forms of a legal entity. The innovation department of one host organisation expressed concern regarding finding a viable governance structure of a legal entity because potential issues with packaging of IP. Other hurdles included employment rules associated, again, with host organisations and a high level of inertia when technical problems surfaced to take attention away from the activity. Despite these hurdles, a comprehensive review of the pros and cons for each type of legal entity has been completed and it has now been agreed that the BioExcel legal entity should take the form of a Social Enterprise. The legal entity is to be founded in Sweden. The first steps associated with company formation have already been taken: a firm of lawyers have been engaged to incorporate the legal entity and at the time of writing they are conducting their background checks on the founders.

---

[33] The founders are all committed to contributing an initial investment in the entity, some casual discussions have already been undertaken with potential other investors to explore the likelihood of additional forms of investment being made available. We have had some qualified positive feedback from these discussions. For reasons of commercial confidentiality, we cannot name those third parties.

6 Finance

Here, we briefly consider our financial needs, from investment through costs to income. Initial work on cost estimates and revenue streams was done in the previous funding phase and outlined in the BioExcel-1 deliverable "D5.4 - Governance and Business Plan"[34].

6.1 Investment Sources

Founder shareholders and investors have come forward and follow-on investors are ready and willing to provide funds for growth in an initial funding round which has not yet been associated with any specific date. The founders have committed to establishing an incorporation fund of approximately €5,000 (the final figure depends upon how many founders will finally make the commitment). Follow on funding at significantly greater levels has been agreed in principle with a European network of business angels.

6.2 Cost Structure

We cannot yet discuss actual costs, as the legal entity has not yet been established. However, we can outline our plans for its initial operations as a sustainable entity. Here we take a high-level view based on the four main cost centres found in any organisation: Facilities, Personnel, Delivery, and Tax. With respect to the Facilities and Personal costs the BioExcel CoE is constrained in what resources it can consume beyond those allocated to related tasks in the work plan.

The academic institutes that form the majority of BioExcel-2 partners place strict restrictions on their members of staff regarding their participation in additional activities and the use of institutional resources for non-core tasks. These rules have precluded academic partners from joining the founder group leading the formation of the BioExcel legal entity. The individuals forming the group of founders are mostly from the SME partners only and any declarations made below, regarding their ability to commit time and resources to the formation and development of the legal entity are fully in line with their own commitment, as well as that of their host business, to the success of the BioExcel legal entity. The BioExcel CoE or the legal entity have no plans to directly employ academic staff.

In order to ensure scientific and technological excellence of the provided products and services, the company will invite the principal investigators of the consortium partners to join its scientific advisory board.

Clients of the BioExcel legal entity will be offered services based on their explicit needs. The services will be designed in the legal entity and composed of local and academic inputs (where required). Any academic inputs required for the commercial services sold by the legal entity will be contracted on the basis acceptable to the host institute of the academic service provider. Such contracts, or preliminary MoUs, with BioExcel-2 consortium partner institutions will be established once the legal entity is registered and operational. The consultancy and training business model is notoriously difficult to scale and some of the academic teams already find it impossible to meet the demands for assistance

---

[34] https://zenodo.org/record/3564177

from the industrial community. Through adopting the "train the trainer" model of working, the BioExcel CoE will be able to build capacity to respond to industry demand, while simultaneously reducing the impact on the academic teams.

6.2.1 Facilities

Given the market conditions and scale of initial investment previously described, the BioExcel legal entity will grow organically from a very small and cost-effective kernel, implying a slow-grow approach to market penetration. As such, we already know that dedicated premises are beyond our reach. We, therefore, plan that the BioExcel legal entity will initially form as a virtual organisation with only a registered address. This may be hosted by an accountant we appoint, or by a business hosting company; such issues will be addressed by the founders as they progress through the company formation process. As we anticipate, there will be no physical premises initially, the personnel assigned to the BioExcel legal entity will work from their existing offices through dedicated phone lines and online systems. Whoever hosts the registered address will have the necessary coordinates in order to forward the mail sent to the registered address to the appropriate person in the BioExcel legal entity's virtual office.

Only when the BioExcel legal entity is generating enough income to become independent will the migration from virtual organisation be considered by the shareholders.

6.2.2 Personnel

Again, due to the planned organic growth from a small kernel and low initial investment, we plan to use sweat equity in lieu of paying salaries to BioExcel legal entity personnel. Sweat equity is a system often used in start-ups that compensates the efforts of personnel with equity rather than salary. How this equity will be vested is something that the founding shareholders will agree during company formation. The founders' equity portfolio will be proportional to the effort they put into the business. How this effort is measured will, again, be discussed by the founding shareholders during company formation.

6.2.3 Delivery

Delivery costs for BioExcel services are anticipated to be nominal; with the initial service offer being that of training and consultancy services, we anticipate that travel and subsistence will be factored into the service fees. We do not anticipate any further delivery costs. However, once the BioExcel CoE starts to grow, it could be that other types of service are offered (such as software modification and the development of bespoke packages); then the Delivery cost centre will need to be revisited.

6.2.4 Tax

The BioExcel CoE will maintain accurate records of its transactions and will pay tax as required by the relevant national authority, once the books are signed off by a qualified accountant, the fees for which are a subcategory in the tax cost centre.

6.3   Income Generation

Again, without real data, the following discussion is based on our plans, which are backed up and based on the preliminary market modelling and initial user needs activities we have carried out.

6.3.1   Market Opportunity

Even the worst-case scenario of the internal market analysis shows that a small market opportunity (approx. $160K p/a now, rising to $320K p/a in 2023 and $500K p/a in 2028) does exist. If successfully exploited, such a level could provide initial support for bootstrapping operations of the legal entity.

6.3.2   Revenue Streams

A convenient correspondence arises when considering the user needs and the initial service offer to be extracted from the BioExcel portfolio. Both have identified training and workflow consultancy as being in demand and also our strongest initial offer. Through the "train-the-trainer" model, we anticipate that our initial income streams will emerge here, and we will be modelling these in detail during the forthcoming period of the project.

Given the correspondence noted above, it seems clear that the initial business model for the BioExcel legal entity is that of offering "support services" to entities engaged in biomolecular research, whether they be in industry or in academia.

Given the positive outcome of the worst case forecast in the modelling carried out and discussed above, there is an opportunity to bootstrap the legal entity's initial operations through addressing the needs of this limited initial market.

7   Next Steps

As a preliminary exercise in modelling the BioExcel market opportunity, it is not appropriate to attempt to draw conclusions here. However, it is appropriate to note some headlines:
- The general understanding of market dynamics at the global level is still at an immature stage of development.
    - We believe this is because the BioExcel market domain is relatively new.
- Internal worse case market modelling reveals that a BioExcel CoE market opportunity does exist.
    - Market modelling efforts will continue and more refined models will be developed to understand this opportunity more clearly.
- BioExcel tools compare well with commercial tools and, overall, in smaller businesses, they are favoured equally with the commercial tools. This sub-domain is where our main market opportunity lays.
    - Some distortion seems to be apparent in the results analysis of our user survey. The causes of this distortion will be avoided in future surveys.

- Distortion through low volume of respondents will be addressed through increasing the profile of the survey and changing the way the survey is delivered.
- Distortion through incorrect question answering will be addressed through addressing the way questions are presented and applying tighter controls over how questions can be answered.
  - BioExcel CoE decisions will be based on opportunity data that are not distorted.
- A legal entity, to oversee sustainable operations, is in the process of being formed.
  - Resources beyond those derived from the project are ready to be committed by commercial partners.

Our efforts in the forthcoming period will focus on developing a more mature understanding of the BioExcel market opportunity and responding formally and operationally to this opportunity. This preliminary business model will be developed into a comprehensive business plan, which is to be developed in two phases. The preliminary business plan (D5.4) will be delivered in PM18 and the final business plan (D5.5) will be delivered in PM30. During this same period, we expect to have launched the BioExcel legal entity and to report on the initial practical aspects associated with market operations.

List of Annexes
    Annex A: Industrial User Survey Questions

Annex A: Industrial User Survey Questions

BioExcel Industry User Survey

Introduction

BioExcel has for many years been the leading European entity providing tools and services to the computational biomolecular academic research community. BioExcel is now seeking to make its products and services attractive to computational biomolecular modelers in industry. This survey is designed to gain insights, from users working in industry, about:
1. **The scientific challenges they may face in computational biomolecular modelling and simulation.**
2. **The technical challenges they may face in the usage of tools and services they rely on when engaged in this kind of work.**

We are collecting this information in order to better understand how our existing and planned tools and services might be able to help industrial users.

1. What size company do you work for?
>Micro enterprise: 10 or fewer employees.
>Small enterprise: 10-49 employees.
>Medium-sized enterprise: 50-249 employees.
>Large enterprise: 250 or more employees.

2. Is the company you work for a multinational enterprise?
(A multinational enterprise is simply one that has a business presence in more than one country)
>Yes
>No

3. Does the company you work for have a dedicated modelling group?
>Yes
>No
>If the modelling is contracted out, please tell us which company performs the modelling for you.
>>Open question. A free text box was provided

4. What business sectors does the company you work for operate within?
>Pharmaceuticals
>Biotechnology
>Chemicals (commodity / speciality / fine)
>Petrochemicals
>Agricultural (e.g. crop protection)
>Food and beverage products
>Cosmetics (including e.g. fragrances & flavour companies)
>Other sector (please specify)
>>Open question. A free text box was provided

5. What are the most common challenges you currently face in computational molecular modelling and simulation?

       Virtual screening for lead discovery

       Antibody engineering and optimisation

       Kinetics of interactions

       Dynamics of molecular complexes

       Molecular determinants of QSAR

       Virtual generative chemistry for lead discovery

       Hit expansion for lead discovery

       Precise binding energy prediction in lead optimisation

       Off-target molecular modelling

       Other challenges not listed here

           Open question. A free text box was provided

6. Below we have listed a number of tools and services that are commonly used to address the challenges of Q5. Please indicate here your satisfaction with them. You can tell us about tools and services we have not anticipated in the final three "Other" selections.

(The respondents were asked to score the tools against the following criteria)

       5. All of my needs are met

       4. Most of my needs are met

       3. Some of my needs are met

       2. Many of my needs are not met

       1. My needs are not met at all

(The list of tools the respondents were asked to score)

       GROMACS local installation

       PMX as an online service

       HADDOCK local installation

       HADDOCK as an online service

       Workflows (e.g. KNIME) consulting at your location

       Workflows (e.g. KNIME) training as an online service

       Face to Face training sessions at your location

       Training as an online service

       Schrödinger platform

       MOE from Chemical Computing Group

       Discovery Suite by Biovia 3DS

       Other Tool/Service (x3)

If you have told us that any of the named tools or services listed above do not meet your needs (fully or partially), please name the Tool/Service and briefly tell us which of your needs are not met and why. If you are telling us anything about the final three "Other Tools/Services", please name these in the order you entered them in the list

Open question. A free text box was provided

7. What features or support might justify your use of different tools or services?

Faster predictions
Greater scale of prediction
Higher quality of prediction
Improved interoperability and standards
Improved user experience
Better technical support
Better tutorials and documentation
Better training on usage
Application consulting
Lower cost (time and money)
Reduce dependency on one provider
Adopting HPC solutions (including GPUs)

Please tell us about any other additional functionality, features or support, not listed above, that you would like to use.
Open question. A free text box was provided

8. Would your organisation be prepared to pay for these enterprise quality features and/or professional support?

Yes
No

9. You answered No to the previous question. Please share your top reason(s) for not wanting such services.

Open question. A free text box was provided

NB When you have completed the answer, please skip Q10 and go directly to Q11

10. You answered Maybe to the previous question. What conditions (for and against) are required to make a purchase decision?

Open question. A free text box was provided

11. We maintain an Industry User Group. The user group is designed to help us deliver the tools and services you actually want. It also allows us to follow-up on this survey and to alert you of, and engage you in, forthcoming new developments from BioExcel and its partners.

If you would like to join our Industry User Group, please provide us with your contact details.

NB We will store your personal data in full accordance with the requirements of GDPR. We will not use your personal data for any purpose

other than administering your own participation in the user group. You are able to withdraw from the user group at any time and if you do this, we will immediately delete your personal data from our records. Finally, the data collected here, and held securely by us, is not associated with the survey data. You cannot be identified from your participation in the survey.

    Name
    Company
    Email address