

H2020 EINFRA-5-2015



Center of Excellence for Computational Biomolecular Research

www.bioexcel.eu

Project Number 675728

D4.6 – Final report on dissemination and training

WP4: Dissemination and Training



Copyright © 2015-2018 The partners of the BioExcel Consortium



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

Document Information

Deliverable Number	D4.6
Deliverable Name	Final report on dissemination and training
Due Date	2018-12-31 (PM38)
Deliverable Lead	EMBL-EBI
Authors	Vera Matser (EMBL-EBI), Michelle Mendonca (EMBL-EBI), Rossen Apostolov (KTH)
Keywords	Training, Dissemination, Impact
WP	WP4
Nature	Report
Dissemination Level	Public
Final Version Date	2018-12-19
Reviewed by	Alexandre Bonvin (UU), Bert de Groot (MPG), Holger Noack (FT), Ian Harrow (IHC), Emiliano Ippoliti (Juelich), Erik Lindahl (KTH)
MGT Board Approval	2018-12-19

Document History

Partner	Date	Comments	Version
EMBL-EBI	2018-12-07	First draft sent to project partners	0.1
EMBL-EBI	2018-12-12	Second draft sent to reviewers and EB	0.2
EMBL-EBI	2018-12-14	Third draft incorporating comments from EB and formatting changes; submitted to PMB	0.3
EMBL-EBI	2018-12-19	Fourth draft, integrating comments from EB and PMB	0.4

Executive Summary

This document describes the dissemination, collaboration and training activities that have been completed during the second period of the BioExcel Centre of Excellence and provides an estimate of the impact that has been achieved by these activities. The dissemination section presents information about the project website, social media channels, dissemination events and publications.

The website has been the main medium for collecting and disseminating information. It attracted on average 1000 visitors/month, 64000 views during the reporting period. The general BioExcel mailing list now has close to 1000 subscribers, while the different Interest Groups (IGs) have between 50-90. Our twitter feed is another major outreach channel with close to 1000 followers and over 1100 tweets to date (top performer among the CoEs).

During the second reporting period, BioExcel has contributed to over 40 dissemination and training events, reaching an estimated 24, 683 people for dissemination and 517 for training. 20 publications have been co-authored by BioExcel personnel or acknowledge our support and services. We present an overview of the Memorandums of Understanding (9 finalized and 2 planned) that have been signed with key EU and US organizations and the impact of the collaborations. We provide a brief update on the activities of the Training Interest Group, the activities during the second period have centered on spreading the word of the methodology used for the BioExcel training programme by working with specific initiatives such as the IMAbs consortium and the ELIXIR implementation study “Developing learning paths for users of ELIXIR services”.

For the face-to-face training events that were completed during the second period, we discuss the event structure and the post-course feedback. The three BioExcel training events, reported on in this deliverable, have been rated as "Good" or "Excellent" by, respectively, 91%, 94% and 100% of the participants. We present the results of the training impact assessment using long-term feedback surveys at 6-12 months and 2 years after the course. The vast majority of our participants (84%) teach others the skills they have learnt, thereby increasing the impact of our training programme and we have been able to facilitate new collaborations.

Joint training events and workshops have achieved a big impact of strengthening the community. The BioExcel/PRACE spring school for the first time brought together core developers of the most popular, open source Molecular Dynamic (MD) codes (GROMACS, AMBER, NAMD) and Visual Molecular Dynamics (VMD). The workshops during PASC18 (Basel, July 2018), Workflows Workshop (Barcelona, December 2018) and Sharing Data for Molecular Simulations (Stockholm, November 2018) were also instrumental in making close ties between European and US researchers, and follow-up events have been planned.

Training activities have been also identified as crucial for the long-term sustainability of the centre and included in the portfolio of services.

Contents

1. GDPR.....	6
2. DISSEMINATION	6
WEBSITE.....	6
MAILING LISTS AND NEWSLETTERS	8
SOCIAL MEDIA.....	9
<i>Twitter</i>	9
<i>LinkedIn</i>	11
<i>Impact of social media</i>	12
<i>Future strategy</i>	13
CONFERENCES, DISSEMINATION AND TRAINING EVENTS	14
PUBLICATIONS.....	18
CONCLUSION ABOUT REACH/IMPACT	20
3. COLLABORATIONS.....	21
MEMORANDUMS OF UNDERSTANDING.....	21
TRAINING INTEREST GROUP	26
4. TRAINING	27
TRAINING EVENTS.....	27
<i>Foundation skills for HPC in computational biomolecular research - BioExcel Summer School</i>	28
<i>Hands-on Introduction to HPC for Life Scientists (Joint BioExcel and PRACE)</i>	31
<i>BioExcel Summer School on Biomolecular Simulations 2018</i>	35
<i>Travel grants</i>	40
IMPACT ASSESSMENT.....	42
<i>Response rate long-term feedback surveys</i>	42
<i>Long-term feedback survey 6 - 12 month</i>	43
<i>Long-term feedback survey 2 years</i>	46
<i>Impact assessment conclusion</i>	46
SUSTAINABILITY.....	46
<i>Knowledge Resource Center</i>	46
<i>Training as a sustainable product</i>	52
5. KPI.....	53
6. CONCLUSION & FUTURE WORK.....	53

1. GDPR

In May 2018 came into effect the new European General Data Protection Regulation ([GDPR](#)) which improves the transparency and data privacy rights of individuals. It underpins our interactions with the community such as communications, surveys, event organization etc. We at BioExcel support and implement compliance with the guidelines. We handle user data responsibly and don't share it with third parties. Our Privacy Policy is fully explained on a dedicated webpage¹. All subscribers of our mailing list have been notified about the terms of service and policy.

2. Dissemination

Here we present information about the project website, social media channels, dissemination events and publications. We assess their reach and impact in light of the main objectives of the centre.

Website

The website is the main medium for collecting, sharing and promoting the activities and results of the center. It's content, both in terms of quantity and categories has grown steadily over the project duration. In **Error! Reference source not found.** we present statistics about the general content, as well as comparative statistics since the last reporting period [PM18].

Table 1: Web portal content statistics

	April 1 2016 - March 31 2017	April 1 2017 - November 30 2018
Number of static content pages	35	34
Number of blog posts	54	118
Number of events	28	55
Pageviews	27991	64798
Sessions	14053	34023
Unique visitors	8996	21270
New users (%)	63	84

¹ <https://bioexcel.eu/about/privacy-policy>

D4.6 – Final report on dissemination and training

Returning Users (%)	36	15
Average session duration (mins)	1:48	1:58
Bounce rate (%)	58.7	63.2
Most users per month	1120 (March 2017)	1899 (March 2018)
Least users per month	522 (December 2016)	812 (December 2017)
Average users per month	750	1180
Top 5 User Locations (% of sessions)	UK (17.3), USA (13.2), Spain (10.7), Germany (6.3), India (6.0)	USA (21), UK (10), India (9), Germany (6), Spain (5)

In **Error! Reference source not found.** we list the 10 most popular pages since the last reporting period (April 2017):

Table 2. Popularity of website pages.

10 most popular pages	% of total page views
Homepage	22
Events	17
Software	7
Categories	6
About	6
Research	3
Services	2.5
News	2.5
Community	2.3
Contact	1.2

As can be seen from the statistics above, the website continues to grow considerably. It is very rich in volume and type of content, which although very positive has presented another challenge - how can we best structure the presentation of content and make it more accessible to different audiences visiting the website? E.g. we have academic users of the software interested in the

D4.6 – Final report on dissemination and training

applications, tools and relevant training opportunities; industrial user to whom we'd like to highlight our commercial offerings; government bodies and funding agencies who'd like to see the impact generated from our work. It is difficult to bring out different aspects without losing some of visitors' interest. For that, we are working on a major restructuring of the front page and menu layouts which we believe will help with making the content easier to discover by the different groups. Work is underway and we expect to launch the new version at the beginning of 2019.

Mailing lists and newsletters

Mailing lists and newsletters have been extensively used for communication with the wider audience. Visitors to the center website are encouraged to subscribe to the general BioExcel mailing list. This mailing list is used to keep users informed about upcoming events, webinars, relevant new material, success stories etc. via mostly monthly newsletters. Registrants to the BioExcel educational webinar series are encouraged to join this mailing list as well. Archives of the newsletters are available at <http://bioexcel.eu/about/newsletters>.

In addition to the general mailing list, we have separate lists of people who have expressed interest in joining the different Interest Groups (IGs). Currently we have the following lists and number of subscribers as of November 30st 2018 (in brackets we have the measure at the previous reporting period Apr 2017):

- BioExcel General Mailing list - 973 (435)
- Entry Level Users IG - 55 (24)
- Free Energy IG - 82 (31)
- Hybrid Methods IG - 79 (25)
- Industry IG - 48 (26)
- Integrative Modelling IG - 79 (33)
- Integrative Modelling webinar series - 88 (70)
- Training IG - 49 (20)
- Workflows - 55 (newly created)

We have reconsidered the interest groups as means of engagement with the specific audiences. It has been challenging to market to the Members yet another activity, while requiring considerable effort by the IG leaders. One drawback of the current interactions is that the mailing lists are for announcements only and do not provide for an easy feedback to the rest of the group. Instead, members were encouraged to use the forums at <http://ask.bioexcel.eu>, but so far, we have not been successful at stimulating activity in the IG-specific forum pages, while the software specific forum pages show high activity in various cases.

In order to address the current state, we are planning on the following alternative actions. IG members will continue be part of the general BioExcel mailing list but most of the IGs (specifically, the IG mailing lists) will be discontinued as such. Instead, we will keep the respective categories as "areas of interest". Website

D4.6 – Final report on dissemination and training

visitors will then be able to select one or more of these areas upon subscription and receive additional information regarding e.g. events, webinars etc. related to the topic.

We should clarify that even though the IG mailing lists are discontinued as such, many of the core activities in the respective areas will continue in future.

Social media

A multi-channel social media strategy is essential for an effective marketing campaign depending on what your interests are and where your audience lies. Over the last year, BioExcel has developed its presence on two social media channels, i.e. Twitter and LinkedIn to drive brand awareness and promote opportunities in the sector. The metrics used in this report to analyze our performance are explained below briefly.

Impressions is the total number of times your content is visible on people's timeline or page.

Engagements is the total number of times a user has interacted with the post, while the engagement rate is the number of engagements divided by impressions. Followers can engage with the content in the form of a retweet, like, mention or direct message. A good engagement rate generally occurs between 1% - 3.3%, which we have used to benchmark results.

Followers can be categorised into followers and organic audience. Followers include all accounts that subscribe to your tweets and organic audience include accounts that you actually reach.

Twitter

Data obtained through Twitter analytics provided insights into the BioExcel account. The social media strategy mainly revolved around providing value to our audience. In order to achieve this, we promoted a variety of content such as blogs, events and latest research in high performance computing and molecular simulations. We also increased the frequency of posting to 3-4 times per working day and engagement with followers.

Growth in Followers

Through the last year, the BioExcel Twitter account has seen a steady growth of followers as seen in Figure 1. As of 30/11/2018, the account (@BioExcelCoE) has 938 followers, which is approximately 3 times higher than the number reported last year (i.e. 225).

D4.6 – Final report on dissemination and training

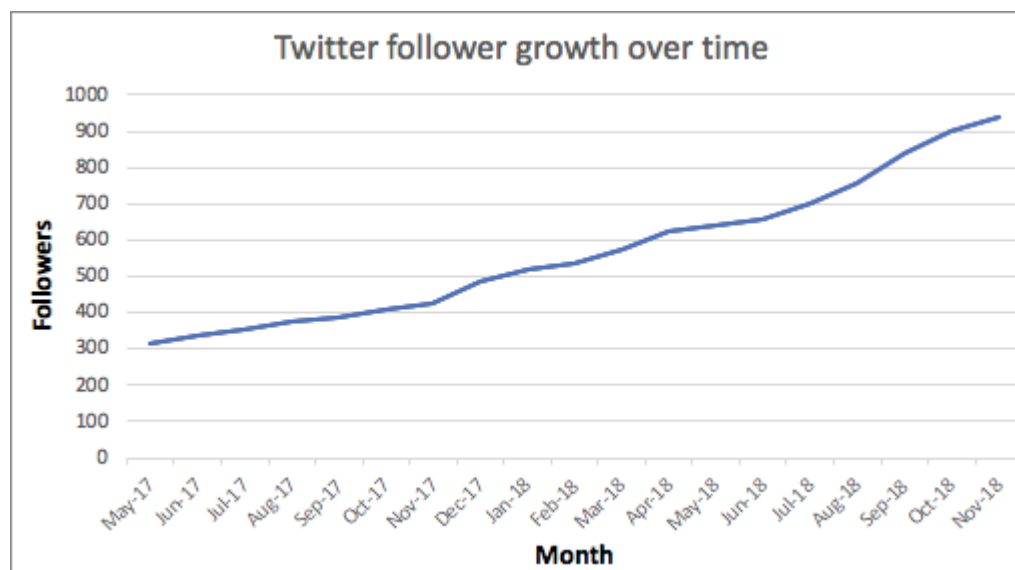


Figure 1: Graph depicting increase in BioExcel twitter followers.

Compared to other Centres of Excellence, BioExcel is leading the board with the highest number of followers and tweets as seen in **Error! Reference source not found.**

Table 3. Number of twitter followers of EU Centres of Excellence.

Twitter Account	Followers	Tweets
@BioExcelCoE	938	1,115
@POP_HPC	499	402
@CompBioMed	527	519
@MaX CoE	492	665
@esiwace	192	362
@ECAM2020	233	175
@NoMaDCoE	587	918
@CoeGSS	323	511

To continue providing value-based content, it is vital to understand our target audience which is primarily broken down by gender and country.

Gender

The gender split for BioExcel followers and organic audience is shown below in Figure 2. The statistics show a huge gap between male and female followers which is expected as the area of high performance computing is largely known to be male

D4.6 – Final report on dissemination and training

dominated. However, we recognize this issue and to counter this effect, we have been engaging online with prominent accounts such as Women in HPC (@women_in_hpc) and promoting research by female scientists.

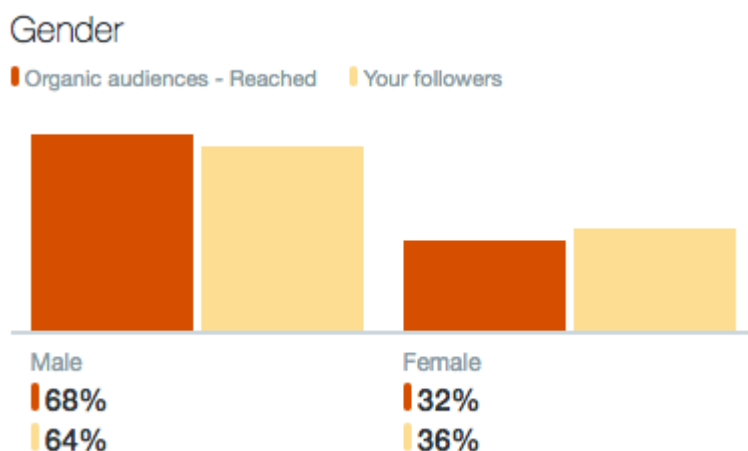


Figure 2. Gender split across @BioExcelCoE followers and organic audience.

Country

Demographic information in Figure 3 shows United States, United Kingdom and Spain dominate the top 3 countries of our followers and organic audience. It is interesting to note almost 20% of our audience is from the United States which signifies our expanding reach.

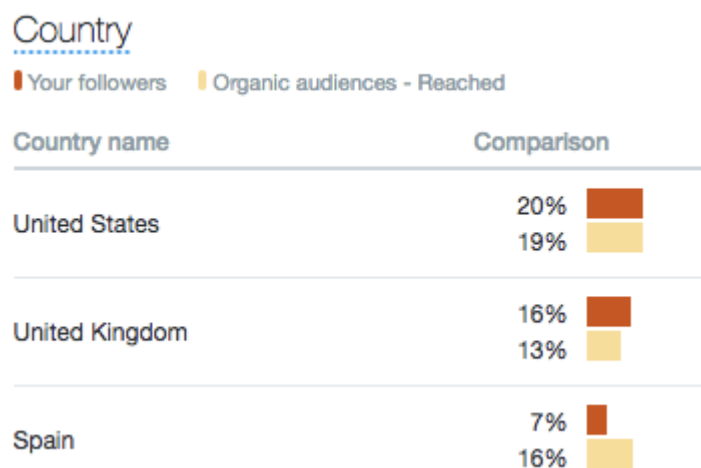


Figure 3. Country of followers and organic audience

LinkedIn

LinkedIn is a great tool to create a meaningful network and build credibility. The BioExcel LinkedIn account has 167 followers and 16 registered employees (as of 30/11/2018). For the purpose of this report, we will look at two primary metrics – Job function and Seniority. The top 2 fields represented here are Research

D4.6 – Final report on dissemination and training

(32.86%) and Education (14.29%) which correlates with the interest groups we actively engage with (Figure 4).

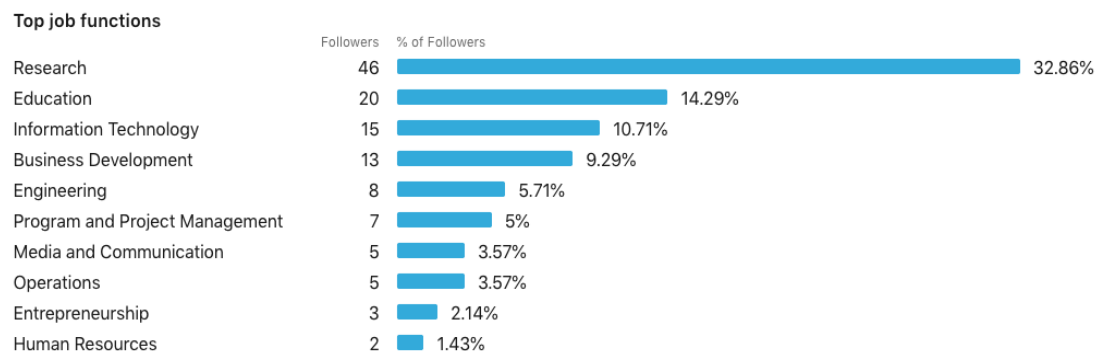


Figure 4. LinkedIn analytics of job functions across followers.

Based on experience level, 39.31% of our followers are entry level closely followed closely by 37.93% senior level (Figure 5).

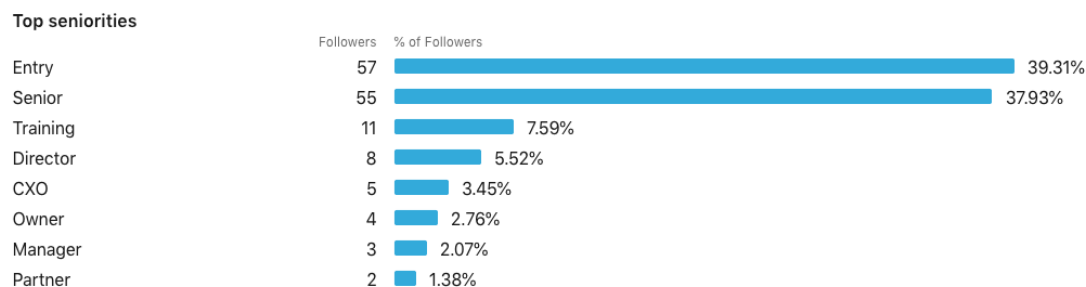


Figure 5. LinkedIn analytics of seniority across followers.

Impact of social media

BioExcel has been successfully able to harness the power of social media marketing displayed through the examples below.

1) 2018 European Conference on Computational Biology (ECCB)

The month of September 2018 had seen the highest number of new followers gained, majorly attributed to BioExcel's presence at ECCB 2018 in Athens. Having a strong social media presence at conferences and boosting online interaction using relevant hashtags allowed us to promote the BioExcel brand and encouraged attendees to visit our booth (Figure 6).

D4.6 – Final report on dissemination and training

Top Tweet earned 4,088 impressions

Booth 2 almost ready - Come talk to us tomorrow about [#training](#) [#software](#) development [#moleculardynamics](#), and of course have a play with the [#virtualreality](#) kit at [#eccb18](#) pic.twitter.com/SIZgMiT4UI

Figure 6. Top performing tweet in September 2018.

2) Workshop on "Sharing Data from Molecular Simulations"

One of our social media strategies was to increase online engagement with followers which was successfully achieved through this event. Although the event was based in Sweden, live tweeting and posting questions online allowed people from all across the world to contribute to the discussion. Session information was also shared with images (Figure 7) to boost engagement and reach of the post. The statistics revealed an engagement rate of 1.3% for the month of November 2018, the highest rate for BioExcel so far.

Top mention earned 239 engagements



matthieu Chavent
@Matth_Chavent · Nov 27

thanks to all the participants it was a wonderful meeting! [#SDMS18](#)
[@BioExcelCoE](#)
[@the_mabraham](#) [@jbarnoud](#) [@amijbonvin](#)
[@jchodera](#) [@karmecon](#) [@DelemotteLab](#)
[@hgrubmu](#) [@eriklindahl](#) [@NMRlipids](#)
[@dga_smith](#) [@pstansfeld](#) [@j0kaso](#)
[@mika_trellet](#) [@chryswoods](#)
pic.twitter.com/Ocbkr30vbe



↳ 2 ↳ 5 ♥ 23

[View Tweet](#)

Top Tweet earned 6,258 impressions

What are the best practices to share data from [#molecular](#) simulations? Follow the hashtag [#SDMS18](#) on 25-27 Nov to participate in the conversation with [@Matth_Chavent](#), [@dga_smith](#) & [@DelemotteLab](#). More info on the [#workshop](#) here: bioexcel.eu/events/worksho...

↳ 19 ♥ 25

Figure 7. Top performing tweets promoting the workshop.

Future strategy

D4.6 – Final report on dissemination and training

In the last year, twitter has been BioExcel’s primary social media channel and seen a considerable growth. In the next phase, we plan to implement the following actions to improve our social media reach.

- 1) Increase the number of posts on LinkedIn allowing us to expand our reach and grow the audience base
- 2) Prepare a list of influencers within the HPC and biomolecular research community to increase our visibility in countries with limited reach
- 3) Analyse metrics to understand benefits of possibly introducing new media channels such as Facebook.

Conferences, dissemination and training events

Since May 2017, the project has actively participated in 40 events featuring BioExcel which are listed below in Table 4. This table covers BioExcel contributions to international and national conferences, knowledge exchange workshops with BioExcel contributions, BioExcel training events, BioExcel partner training events where BioExcel resources were presented, and any other relevant events. 21 events were labelled as dissemination, 16 events combined dissemination and 3 events were labelled as Training. The total number of people reached through dissemination was 24,683 and 517 for training.

Table 4. List of Dissemination and Training Events. Type indicates D=dissemination, T= training, T/D= training and dissemination.

No	Date	Event Name	Location	Attendees	Type
1	31 May 2017	Free Energy Calculations from Molecular Simulation: Applications in Life and Medical Sciences*	London, UK	40	D
2	1 June 2017	Preparing for PRACE Exascale Systems	FZJ, Germany	20	D
3	3 -7 July 2017	Foundation skills for HPC in computational biomolecular research - BioExcel Summer School*	Hinxton, UK	25	T
4	29 November - 1 December 2017	“Hands-on Introduction to HPC for Life Scientists” in collaboration with PRACE*	Edinburgh, UK	17	T

D4.6 – Final report on dissemination and training

5	10 January 2018	MeetU symposium	Paris, France	20	D
6	15-19 January 2018	FIRST project workshop	Hanoi, Vietnam	30	T/D
7	30 January 2018	From Virtual Reality to Immersive Analytics in Bioinformatics	Burlingame, California	100	D
8	17-21 February 2018	62nd Annual Meeting, Biophysical Society	San Francisco, California	6000	D
9	19-23 February 2018	INSTRUCT practical course on “Advanced methods for the integration of diverse structural data with NMR data	Florence, Italy	20	T/D
10	8-9 March 2018	IT Transformation and cloud content management for life sciences	Berlin, Germany	100	D
11	16-23 March 2018	ISGC2018	Taipei, Taiwan	200	D
12	9 May 2018	HADDOCK workshop	Vienna, Austria	20	T
13	15 May 2018	Pint of Sciences 2018	Utrecht, The Netherlands	50	D
14	24 May 2018	SciLifeLab Biomolecular Simulation Workshop*	Uppsala, Sweden	30	T/D
15	28 May - June 1 2018	DynaMoPPI : Dynamique et Modulation des Interactions Protéine-Protéine	Nantes, France	20	T/D
16	29 - 31 May 2018	PRACEdays18	Ljubljana, Slovenia	100	D

D4.6 – Final report on dissemination and training

17	3-8 June 2018	NMRFAM structure determination workshop	Madison WI, USA	30	T/D
18	11-14 June 2018	NMBOX summer school	Farmington CT, USA	30	T/D
19	2-6 July 2018	EMBO practical course	Barcelona, Spain	25	T/D
20	6 - 10 July 2018	ISMB 2018	Chicago, IL, USA	2000	D
21	16-21 July 2018	Course on integrative structural biology	Paris, France	25	T/D
22	13-17 August 2018	NMR Practical Course 2018 Multidimensional NMR in Structural Biology	Joachimsthal, Germany	40	T/D
23	13-17 August 2018	School: computer simulations of biological membranes & free energy calculations of biomolecular systems*	Bogota, Colombia	30	T/D
24	19-24 August 2018	27th ICMRBS	Dublin, Ireland	100	D
25	3-7 September 2018	Structural Bioinformatics course*	Hinxton, UK	25	T/D
26	8 September 2018	SIG 2018*	Athens, Greece	33	D
27	8 - 12 September 2018	ECCB 2018*	Athens, Greece	1000	D
28	11 September 2018	Helix Nebula SciCloud training meeting	Amsterdam, The Netherlands	50	D

D4.6 – Final report on dissemination and training

29	11 October 2018	From High Performance Computing to Exascale: the research challenges and new applications	Milan, Italy	30	D
30	16-19 October 2018	CECAM workshop on Protein-peptide interactions	Paris, France	35	T/D
31	17-18 October 2018	GROMACS 2018 developer/power-user workshop*	Goettingen, Germany	30	T/D
32	5-6 November 2018	BioExcel/PDBe Hackathon*	Hinxton, UK	5	T/D
33	8-10 November 2018	Modeling of Protein Interaction meeting	Kansas, USA	40	T/D
34	11-16 November 2018	SC18	Dallas, Texas, US	14000	D
35	16 November 2018	Netherlands Society on Biomolecular Modelling*	Utrecht, The Netherlands	30	D
36	25 - 27 November 2018	Sharing Data from Molecular Simulations*	Stockholm, Sweden	15	D
37	28 November 2018	International Conference on artificial intelligence and robotics for industrial applications	Putrajaya, Malaysia	100	D
38	29 November 2018	Minisymposium: Biomolecular simulation in drug discovery*	Solna, Sweden	20	D
39	29-30 November 2018	Molecular & Computational Biology Symposium 2018	Dublin, Ireland	100	D

D4.6 – Final report on dissemination and training

40	11- 14 December 2018	BioExcel/MolSSI Workshop on Workflows in Biomolecular Simulations	Barcelona, Spain	20	T/D
----	----------------------------	---	---------------------	----	-----

Publications

Here we present the publications of research work that has been supported by BioExcel during the second reporting period (table 5).

Table 5. Publications supported by BioExcel.

** Case studies and recommendations in this publication are based on a 2017 workshop which was sponsored by BioExcel “Centre Européen de Calcul Atomique et Moléculaire (CECAM), Ion Transport from Physics to Physiology: the Missing Rungs in the Ladder”.*

Publication type	Authors	Year	Title	Journal Name	DOI
Journal article	S. Möller et. al	2017	Robust cross-platform workflows: how technical and scientific communities collaborate to develop, test and share best practices for data analysis	Data Science and Engineering	https://doi.org/10.1007/s41019-017-0050-4
Journal article	V. Gapsys et al	2017	Alchemical free energy calculations for nucleotide mutations in protein-DNA complexes	Journal of Chemical Theory and Computation	https://doi.org/10.1021/acs.jctc.7b00849
Journal article	I.S. Moreira, et. al	2017	SpotOn: High Accuracy Identification of Protein-Protein Interface Hot-Spots.	Nature Scientific Reports	http://dx.doi.org/10.1038/s41598-017-08321-2
Journal article	E. Karaca et. al	2017	M3: An Integrative Framework for Structure Determination of Molecular Machines.	Nature Methods	http://dx.doi.org/10.1038/nmeth.4392
Journal article	V. López-Ferrando et. al	2017	PMut: a web-based tool for the annotation of pathological variants on proteins, 2017 update.	Nucleic Acids Research	https://doi.org/10.1093/nar/gkx313
Newsletter	R. Apostolov	2017	BioExcel Assists Release of Workflow Manager for Large Genetic Studies	PDC Newsletter (bi-yearly)	https://www.kth.se/polopolyfs/1.740478!/Newsletter201

D4.6 – Final report on dissemination and training

					7-1-final-res.pdf
Newsletter	R. Apostolov	2017	BioExcel Update	PDC Newsletter (bi-yearly)	https://www.kth.se/polopolyfs/1.784572!/Newsletter2017-2-final-spreads-res.pdf
Journal article	Z. Kurkcuoglu et. al	2018	Performance of HADDOCK and a simple contact-based protein–ligand binding affinity predictor in the D3R Grand Challenge 2	Journal of Computer-Aided Molecular Design	http://dx.doi.org/10.1007/s10822-017-0049-y
Journal article	A. Rubio-Cosials et. al	2018	Protein Flexibility and Synergy of HMG Domains Underlie U-Turn Bending of DNA by TFAM in Solution.	Biophysical Journal	http://doi.org/10.1016/j.bpj.2017.11.3743
Journal article	A. Garcia-Lopez et. al	2018	Targeting RNA structure in SMN2 reverses spinal muscular atrophy molecular phenotypes.	Nature Communications	http://doi.org/10.1038/s41467-018-04110-1
Journal article	K. Elez et. al	2018	Distinguishing crystallographic from biological interfaces in protein complexes: Role of intermolecular contacts and energetics for classification	BMC Bioinformatics	https://doi.org/10.1186/s12859-018-2414-9
Journal article	A. Basciu et. al	2018	An enhanced-sampling MD-based protocol for molecular docking.	BioRxiv	https://doi.org/10.1101/434092
Journal article	A. Vangone et. al	2018	Large-scale prediction of binding affinity in protein-small ligand complexes: the PRODIGY-LIG web server.	Bioinformatics	https://doi.org/10.1093/bioinformatics/bty816
Journal article	P.I. Koukos et. al	2018	Protein-ligand pose and affinity prediction. Lessons from D3R Grand Challenge 3.	Journal of Computer-Aided Molecular Design	https://doi.org/10.1007/s10822-018-0148-4
Journal article	A.M.J.J. Bonvin et. al	2018	Defining distance restraints in HADDOCK	Nature Protocols	https://doi.org/10.1038/s41596-018-0017-6

D4.6 – Final report on dissemination and training

White paper	M.J. Abraham et. al	2018	BioExcel Whitepaper on Scientific Software Development.	Zenodo	http://doi.org/10.5281/zenodo.1194634
Journal article	T. Bastys et. al	2018	Consistent prediction of mutation effect on drug binding in HIV-1 protease using alchemical calculations	Journal of Chemical Theory and Computation	http://doi.org/10.1021/acs.jctc.7b01109
Journal article	R. Briones et. al	2018	GROmaps: a GROMACS-based toolset to analyse density maps derived from molecular dynamics simulations	Biophysical Journal	http://doi.org/10.1016/j.bpj.2018.11.3126
Journal article	P.I. Koukos et. al	2018	A membrane protein complex docking benchmark.	Journal of Molecular Biology	https://doi.org/10.1016/j.jmb.2018.11.005
Journal article	C. Geng et al	2018	iSEE: Interface Structure, Evolution and Energy-based machine learning predictor of binding affinity changes upon mutations.	Proteins: Struct. Funct. & Bioinformatics Advanced Online Publication	https://doi.org/10.1002/prot.25630
Journal article	V. Lindahl et. al	2018	Permeability and ammonia selectivity in aquaporin TIP2;1: linking structure to function	Scientific Reports	https://doi.org/10.1038/s41598-018-21357-2
Report	E. Lindahl (editor in chief)	2018	The Scientific Case for Computing in Europe	PRACE	ISBN 9789082169492
Journal article*	R.J. Howard et. al	2018	Permeating disciplines: Overcoming barriers between molecular simulations and classical structure-function approaches in biological ion transport.	Biochimica et Biophysica Acta (BBA) - Biomembranes	https://doi.org/10.1016/j.bba.2017.12.013
Newsletter	R. Apostolov	2018	BioExcel Update	PDC Newsletter (bi-yearly)	https://www.kth.se/polopolyfs/1.830220!/Newsletter2018-1-final-lres.pdf

Conclusion about Reach/impact

The web portal has grown into a content rich platform that attracts a steady number of visitors in line with the projected audiences. Initiating conversations on social media helps us grow our brand value and promote our effort to improve biomolecular software packages and best practices. Our Twitter channel has a

D4.6 – Final report on dissemination and training

continuously growing number of followers and supports the popularization of the brand.

Publications in high-impact journals has contributed to raising the profile of the centre and the quality of our work. The core applications continue to be cited on average between 300 and 4000 times/year. Presentations at international and national conferences has been a great platform to network and assess the needs of the community. Such events offered chances on a number of occasions to establish new collaborations for future funding of activities in line with the BioExcel objectives.

In the future, it would be worthwhile for BioExcel to organize our own conference, considering we have built a loyal user base through the years. Initial plans are in place for 2020.

3. Collaborations

Extensive engagement with biomolecular research communities, organizations, pharmaceutical/chemical/food industry companies and hardware vendors continued in line with our outreach objectives.

Memorandums of Understanding

Collaborations are an integral for provision of support and impact within the wider community². The centre has formalized several strategic and collaborative partnerships³ through Memorandums of Understanding (MoUs) with the following organizations:

- Open PHACTS foundations (<http://openphacts.org/>) – MoU for Strategic Partnership signed
- ELIXIR (<http://elixir-europe.org/>) – MoU for Collaborative Partnership signed
- MolSSI (<http://molssi.org/>) - MoU for Strategic Partnership signed
- Vrije Universitet Amsterdam (<http://vu.nl/>) - MoU for Strategic Partnership signed
- HPC-Europa3 (<http://www.hpc-europa.eu/>) - MoU for Strategic Partnership signed
- Scuola Normale Superiore (<http://sns.it/>) - MoU for Strategic Partnership signed
- Multiscale Complex Genomics Virtual Research Environment (MuG VRE, <http://www.multiscalegenomics.eu/MuG/>) - MoU for Strategic Partnership signed
- VI-SEEM VRE (<https://vi-seem.eu/>) - MoU for Strategic Partnership signed
- INSTRUCT (<https://www.structuralbiology.eu/>) – pending final counter-signature for a Strategic Partnership MoU

² <https://bioexcel.eu/community/collaborations>

³ details about the nature of the collaborations is given in “D4.4 - Dissemination Report and Update Plan” - <https://doi.org/10.5281/zenodo.574614>

D4.6 – Final report on dissemination and training

At the time of writing of this document, we have started a discussion for two more MoUs with HECBioSim (<http://www.hecbiosim.ac.uk>) and CCPBioSim (<http://www.ccpbiosim.ac.uk/>).

It should be noted that strategic and collaborative partners have been selected based on expected impact and leverage of joint activities.

Impact

Open PHACTS is a platform that brings together vast amounts of linked data in a single integrated, open infrastructure for researchers to access and query publicly-available pharmacological data. The platform integrates major open data sources (e.g. ChEMBL, UniProt, WikiPathways, SureChEMBL, DrugBank, Disgenet and their associated Chemistry) on compounds, patents, diseases, targets and pathways; linked by identifiers, semantic and structural chemistry information, re-exposed as a collection of unified web APIs, workflow components plus tool integrations.

Impact: This was our first official partnership with an external initiative. Our interest was in integrating in our workflows some of the data exposed by Open PHACTS. In addition, the CTO of Open PHACTS, Nick Lynch, is a member of the BioExcel Scientific Advisory Board. As an installable platform Open PHACTS showed to be a deployment challenge for HPC workflows, as it relies on intra-connected containers accessed over the network rather than working as a command line tool. We continue the evaluation of integrating parts of the platform in some custom workflow solutions. In addition, Open PHACTS had already worked on KNIME bindings, and such bindings were flagged as important for industrial partners of BioExcel. Thus, later development of KNIME nodes for the BioExcel Building Blocks were to some extent built on experience from the Open PHACTS community. We have also aired a joint webinar⁴ presented by Nick Lynch.

ELIXIR is one of Europe's leading life science organisations in managing and safeguarding the increasing volume of data being generated by publicly funded life science research. It coordinates, integrates and sustains bioinformatics resources across its member states and enables users in academia and industry to access services that are vital for their research. The collaboration framework between Elixir and BioExcel includes work on the <http://bio.tools> database, adoption of API libraries and workflow specifications, joint efforts towards adoption and performance evaluation of the common workflow language (CWL) in selected workflows, as well as joint webinar series.

Impact: BioExcel is very strongly engaging with ELIXIR and has produced a significant impact in the operations of the organization⁵. This partnership is another example of an excellent opportunity to reach a new organized community, in this case the whole European life science community. Starting by a proposed BioExcel use case⁶ for workflows development best practices connecting the

⁴ <https://bioexcel.eu/webinar-openphacts/>

⁵ <https://bioexcel.eu/bioexcel-centre-of-excellence-in-elixir-all-hands-berlin-2018/>

⁶ <https://drive.google.com/file/d/0B8FFnMih2s3bdDJGWNXYVmwWSE0/view>

D4.6 – Final report on dissemination and training

ELIXIR Tools and Interoperability platforms, and continuing with participations in a series of meetings and conferences, BioExcel has been presented to the whole ELIXIR community, and is now considered as a strong ally for the biomolecular structural field of the life sciences, to the point that BioExcel will be an important part of the new proposed ELIXIR structural community ([3D-BioInfo](#)). BioExcel has also been involved in the [ELIXIR implementation study for CWL](#), which have been improving scalability of CWL workflows on EBI's LSF cluster. This involvement has become a use-case both for our [CWLProv](#) capture of workflow provenance ([10.5281/zenodo.1966881](https://doi.org/10.5281/zenodo.1966881)), [as well](#) as for the collaboration with the US Food and Drug Administration (FDA)-supported [BioCompute Object](#) community effort for regulatory submissions of high-throughput sequencing for personalized medicine ([10.1101/191783](https://doi.org/10.1101/191783)), a project with strong links to industrial pharma.

The Molecular Sciences Software Institute (**MolSSI**) can be seen as a US counter partner organization similar to BioExcel. The Institute supports advances in software infrastructure, education, standards, and best-practices that are needed to enable the molecular science community to open new windows on the next generation of scientific Grand Challenges. Joint activity plans included joint projects involving scientists on both sides, sharing of expertise and/or biomolecular software, organizing joint workshops and training events, as well as webinar series.

Impact: Partnering with MolSSI has been an important and fruitful activity to strengthen the links with leading US researchers and expand the scope and impact of BioExcel expertise worldwide. After the first joint event, a [workshop during PASC18](#), our two organizations joined efforts on 1) development of workflow solutions for HPC (first workshop in Barcelona⁷, Dec. 2018) for gap analysis, and an implementation hackathon follow-up in US in 2019); 2) improvement of data sharing practices for molecular simulations (first workshop in Stockholm⁸, Nov. 2018, and a follow-up planned in 2019); Further plans on additional, large-impact, activities in areas related to free energy calculations, exascale automations and others have been planned. Two joint webinars^{9,10} were also produced and aired.

Within the Division of Molecular and Computational Toxicology at **Vrije Universiteit Amsterdam** (VUA), the Biomolecular Simulation and Modeling team develops, implements and applies advanced in silico models to study and predict protein-ligand binding from (HPC) simulation. The team integrates molecular dynamics and other advanced modeling approaches into efficient heterogeneous workflows for the purpose of binding affinity prediction, protein engineering or other applications. The workflow framework developed by the group of Daan P. Geerke, [MDSudio](#), is a valuable addition to the work at BioExcel. As part of our

⁷ <https://bioexcel.eu/events/bioexcel-molssi-workshop-on-workflows-in-biomolecular-simulations/>

⁸ <https://bioexcel.eu/events/workshop-on-sharing-data-from-molecular-simulations/>

⁹ <https://bioexcel.eu/webinar-molecular-simulation-control-gmxapi-2018-09-19/>

¹⁰ <https://bioexcel.eu/open-force-field-initiative-the-smirnoff-format-and-learned-chemical-perception/>

D4.6 – Final report on dissemination and training

collaboration, the framework was extended with support for our core applications HADDOCK and GROMACS, and a joint webinar¹¹ organized.

Impact: Partnering with the group of Daan Geerke from the Vrije Universiteit Amsterdam was very fruitful in the context of the use case Biomolecular Recognition, for which we were seeking a flexible, yet powerful workflow manager that we could easily be plugged to HADDOCK as a web service. They also contributed to the success of the community forum, pitching in the emerging the challenge of predicting the structure of protein complexes with covalently bound ligands. Another positive outcome of our collaboration was the organisation in September 2017 of a mini symposium entitled "Automation in Biomolecular Simulation and Modeling", which gathered several computational biology teams in the Netherlands and the team of Alan Mark from the University of Queensland in Brisbane, Australia. It created the opportunity to present projects related to the automation of protein modelling using workflows and web services, two core expertises of both VUA and UU teams.

HPC-Europa3 project is based on a programme of visits, in the form of traditional transnational access, with researchers visiting HPC centres and/or scientific hosts who mentor them scientifically and technically. The visitors are funded for travel, accommodation and subsistence. They are provided with an amount of computing time suitable for the approved project. Our two organizations are collaborating on dissemination of relevant information among the respective communities, and exchange of expertise in sustainability development.

Impact: At the time of writing of the deliverable, BioExcel partners have been hosts to 3 visitors by the program, two of which resulted in future collaborations.

The **Scuola Normale Superiore** (SNS) is an Italian public institute for higher education and research. SNS has very active groups in the field of Theoretical and Computational Chemistry linked with several international networks and societies. BioExcel organized joint research work with SNS researchers on further development of multiscale modeling techniques. Our organizations also organized thematic meetings and dedicated schools in the area of biomolecular modeling using hybrid multiscale methods.

Impact: The partnership has led to the establishment of joint workshops between BioExcel and SNS members (e.g. funded by CECAM), where BioExcel codes such as the novel QM/MM interface have been presented. Moreover, some members of SNS actively collaborated in the Hybrid Method Interest Group of BioExcel. Thanks also to the large impact that SNS has in Italy, these initiatives has also led to the interest of several other members of the chemistry and physics Italian community for BioExcel (when the project started, Italy was one of the largest EU country not directly represented in BioExcel) and its Interest Groups, measurable for example by the significant increment of subscriptions to BioExcel activities (socials, IG, newsletters) coming from addresses from Italian institutions.

¹¹ <https://bioexcel.eu/webinar-mdstudio-microservice-based-molecular-dynamics-workflows-2017-10-11/>

D4.6 – Final report on dissemination and training

Multiscale Complex Genomics Virtual Research Environment (**MuG VRE**) supports the expanding 3D/4D genomics community by developing tools to integrate the navigation in genomics data from sequence to 3D/4D chromatin dynamics data. MuG VRE services include: personal user workspace, community-tailored tools and workflows and integrated user support. In addition, MuG offers training activities and community events, with the aim to enhance multidisciplinary communication and contribute to standardization in tools and methods. BioExcel is engaging with MuG researchers on provisioning of tailored solutions such as VMs for the needs of the genomics community, defining a long-term collaborative framework for training, organizing thematic meetings, dedicated schools and educational webinars for increasing the adoption and expertise in biomolecular techniques within relevant genomics communities.

Impact: Partnering with [Multiscale Complex Genomics](#) (MuG) project was an excellent opportunity to reach a recently growing community: 3D/4D genomics [[Nature Genetics, 50, 352–1358 \(2018\)](#)]. BioExcel work, focused on DNA/RNA, protein-nucleic acids complexes and chromatin dynamics simulations and analyses in particular, were presented to the MuG community through a joint MuG-BioExcel [webinar](#) and [workshop](#) (more details [here](#)). The workshop was run just after an important biomolecular simulations meeting organized by the MuG community ([ISQBP](#)), ensuring the outreach range of BioExcel. Reception of both the workshop and the work presented by BioExcel was great, with no places available one month before the planned date, and a good response and interest from the audience.

VI-SEEM VRE provides networking and computational resources, application support and training, in both South East Europe and Eastern Mediterranean. It has supported the European vision of inclusive and smart growth, based on knowledge and innovation, enriching the European Research Area. The project unifies existing e-Infrastructures into an integrated platform to better utilize synergies, for an improved service provision within a unified Virtual Research Environment to be provided to scientific communities of high impact in the combined SEEM region. BioExcel is partnering with VI-SEEM on dissemination of relevant information among the respective communities, promotion of BioExcel events to the VI-SEEM members in particular the Life Science groups, joint training activities including webinars and joint industry outreach activities.

Impact: BioExcel partnered with VI-SEEM for the organization of the SIG BioExcel: 2nd edition meeting during ECCB2018 in Athens. The meeting was very successful and attracted more than 30 attendees. The rich program and high-profile speakers were highly appreciated by the participants as described in their contributed blog posts¹². Future interactions with Vi-SEEM, and related communities, will be important step towards providing expertise and support to countries in south-east Europe.

¹² <https://bioexcel.eu/hear-from-our-2018-sig-meeting-travel-grant-beneficiaries/>

INSTRUCT has established a distributed pan-European infrastructure of high-end technologies for integrated structural biology. The scope includes the traditional techniques of crystallography, NMR and electron microscopy, but aims to encourage the integration of these and other technologies, extending to techniques including cryo-electron tomography, X-ray microscopy and correlative microscopy. INSTRUCT, through its Centres, offers access to its infrastructure for researchers and industrial users and supports this with training programmes and other networking events. A key part of the services offered to users includes computational data services for experimental analyses from structural data. Collaboration activities with BioExcel will be in two main directions: Training (hosting joint training activities in new or specialist computational services or related methods; providing support in the planning, running and dissemination of joint training activities) and Networking (raising awareness among users to the services of the collaborating party, publishing on the websites relevant information about related activities and their promotion). Concrete joint activities will start in 2019.

Training Interest Group

The Training Interest Group was created to aid collaboration with other training professionals, and to interact or collaborates with other training initiatives. According to the recommendations of the EU project review, the activities of the second reporting period have focused more on sharing our expertise in building a structured training programme with other initiatives.

The BioExcel training programme was presented in a talk “Implementing a competency-based training strategy for biomolecular researchers with high computational needs” during the Education session of the ISMB 2018 (26th Conference on Intelligent systems for molecular biology), which took place 6-10th July in Chicago, United States. All sessions were recorded and are available for viewing on YouTube¹³; this recording is a valuable resource for us to share with other organisations that are interested in building a competency-based training programme.

There are two initiatives that we have worked with more closely, the ELIXIR training platform and the iMAbs consortium. The Training Interest group activities are discussed in more detail in D3.5 - Consultancy Modalities and Funding Options, here we provide only a short summary. BioExcel was invited to join the ELIXIR implementation study “Developing learning paths for users of ELIXIR services”¹⁴ as an external collaborator in recognition of the work carried out on competencies and curriculum design. Vera Matser, on behalf of BioExcel, co-organised and delivered the first face-to-face workshop for the implementation study “Applying bioinformatics and data science competency frameworks to ELIXIR Training” on 4-6th April 2018 in Amsterdam. EMBL-EBI is a partner organisations, representing BioExcel, in the iMAbs Innovative Training Networks

¹³

https://www.youtube.com/watch?v=nL6_4xyoCDE&list=PLmX8XnLr6zeGLPYD5q_xC6rz4o7Sf6a_vI&index=184

¹⁴ <https://www.elixir-europe.org/about-us/implementation-studies/learning-paths-2018>

(ITN) proposal, Integrated Modelling of Antibodies “IMAbs”. During the proposal writing stage we have advised IMAbs in building a personalised training programme for Early Stage Researchers using a variant of the BioExcel competency framework. The proposal was not funded but will be re-submitted in 2019; if funded, the work carried out in IMAbs will align well with the planned BioExcel-2 training activities.

4. Training

Here we present an overview of the training activities carried out during the second reporting period of BioExcel. For the BioExcel face-to-face training events, we discuss the event structure and post-course feedback. We present the results of the training impact assessment for all BioExcel training courses using our long-term feedback surveys at 6-12 months and 2 years after the course.

Training events

On average we have organised two main BioExcel training events per year, though there are many more community events or training events at partner sites that BioExcel has contributed to. A full list can be seen in Table 6.

The first three events have been reported on in D4.5 - Training report and updated plan¹⁵ and the last three are described in this document.

Event 1 - BioExcel addressing training needs for advanced simulations in biomolecular research (2016)

Event 2 - Workflow Training for Computational Biomolecular Research (2016)

Event 3 - BioExcel PRACE Spring School - HPC for Life Sciences (2017)

Event 4 - BioExcel Summer School - Foundation skills for HPC in computational biomolecular research (2017)

Event 5 - BioExcel PRACE Hands-on Introduction to HPC for Life Scientists

Event 6 - BioExcel Summer school on Biomolecular simulations (2018)

Table 6. Attendance and Feedback from training events.

Course	Number of attendees	Attendance rate	% feedback	% rated the event Good/Excellent	% would recommend
Event 1	19**	63%	42%	100%	n/a
Event 2	28	93%	79%	95%	91%
Event 3	55	85%	75%	77%	87%
Event 4	25	44%	100%	91%	91%
Event 5	25	72%	89%	94%	94%
Event 6	30	100%	90%*	100%	100%

¹⁵ <https://doi.org/10.5281/zenodo.574620>

Foundation skills for HPC in computational biomolecular research - BioExcel Summer School

The fourth BioExcel training event "Foundation skills for HPC in computational biomolecular research - BioExcel Summer School" took place on 3-7th July 2017 at the European Bioinformatics Institute (EMBL-EBI), Hinxton, UK.

Course overview

The 5-day BioExcel Summer School used project-based learning to empower life scientists to get the most out of using computers. Using example problems and challenges that life-scientists come across in their day-to-day work, e.g. obtaining data, installing software, cleaning data, analysing data; this course helps participants gain the foundational computational skills needed to move into the domain of high-end computing.

Audience

The course was a primer for people who want to start using High Performance Computing (HPC) but find the concept quite daunting and lack the needed computational skills. This summer school was project-based, formal classroom teaching was kept to a minimum. No prerequisites were required for this course, participants from any career stage were encouraged to apply for this course.

Syllabus, tools and resources

The summer school explored concepts such as computer architecture, hardware and software requirements, operating systems, file structure, non-trivial software installation, using Linux commands, automating tasks using the Unix shell, software documentation & best practice.

Learning outcomes

After this course delegates will be able to or have acquired:

- Confidently navigate through the file structure and operating system of the computer.
- Working knowledge of Linux, including basic command line operations
- Obtain & submit data from/to a range of external sources
- Compile, install and test software
- Read and modify scripts

What did this course not do:

- This course did not teach delegates how to programme, though they are much better prepared after this course to start a self-taught course
- This was not a course on HPC computing but did teach delegates the foundational skills and concepts that they need to start learning about HPC computing.

D4.6 – Final report on dissemination and training

Table 7: Programme for "Foundation skills for HPC in computational biomolecular research - BioExcel Summer School" on 3-7th July 2017 at the European Bioinformatics Institute (EMBL-EBI), Hinxton, UK

Time	Topic	Speaker
Day 1 - Monday 3rd July 2017		
10:00 – 11:00	Introduction to EMBL_EBI and BioExcel	Vera Matser, EMBL-EBI
11:00 – 12:00	Course expectations and User stories	All
12:00 – 13:00	Lunch	
13:00 – 15:00	The computational roadmap; from your local machine to clusters and HPC	All, discussion led by Lee Larcombe, Brett McClintock, Mohamed Alibi and Arno Proeme
15:00 – 15:30	Break	
15:30 – 17:30	Introduction to project work (project 0)	Vera Matser Lee Larcombe, NexaSTEM
18:00 – 20:00	Evening meal at Wellcome Trust Conference Centre	
Day 2 - Tuesday 4th July 2017		
09:00 - 09:30	Summary of Day 1 and introduction to concepts of Day 2 using the computational route map	All
09:30 - 12:30	Project 1	
12:30 - 13:30	Lunch	
13:30 - 14:00	Visit to the computing facility	
14:00 - 17:30	Project 2	Lee Larcombe Vera Matser Csaba Halmagyi
18:00 - 20:00	Evening meal (Red Lion, Hinxton)	
Day 3 - Wednesday 5th July 2017		
09:00 - 09:30	Summary of Day 2 and introduction to concepts of Day 3 using the computational route map	
09:30 - 12:30	Project 3	
12:30 - 13:30	Lunch	

D4.6 – Final report on dissemination and training

13:30 - 16:30	Project 4	
16:30 - 21:00	Free evening in Cambridge (bus provided)	
Day 4 - Thursday 6th July 2017		
09:00 - 09:30	Summary of Day 3 and introduction to concepts of Day 4 using the computational route map	
09:30 - 12:30	Project 5	
12:30 - 13:30	Lunch	
13:30 - 17:30	Project 6	
18:00 - 20:00	Evening meal (Wellcome Trust Conference Centre)	
Day 5 - Friday 7th July 2017		
09:00 - 09:30	Summary of Day 4 using the computational route map	
09:30 - 13:00	Introduction to HPC computing	Arno Proeme
13:00 - 14:00	Course feedback and wrap up	Vera Matser
	Lunch vouchers will be provided - End of course	

Conclusion and course feedback

In total 45 people applied for the course (51% female, 49% male), the applicants were a very international group with a high number of applicants from outside Europe (60%). Screening the applications there was a substantial group of people whose motivation did not seem to fit the profile of the course very well (e.g. “want to learn bioinformatics/MD skills”) or were considered too advanced for the course. During the application stage of this event a question was included related to travel grants “If eligible, would you like to be considered for a travel grant?”, a total of 35 applicants replied *Yes* (78%). The course budget was setup to provide financial support for three applicants only. A number of applicants who received an invitation to register had to withdraw due to the (relatively) high costs of attending a 5-day course in Cambridge. With three last minute cancellations/no-shows on the day, the number of people on the course was disappointing, a total of 11 attended the course (64% female, 36% male); this translates to an attendance rate of 44%.

We have asked the participants to fill in our standardised course feedback survey, the response rate on the course feedback survey was 100%. 91% rated the course as *Good* or *Excellent*, 100% of the respondents rated the overall course organisation as *Good* or *Excellent*. When asked about the balance between theoretical and practical content 82% replied *About right*, the remaining 18% felt the course focused too much on practical content. 91% of respondents would recommend the course, with the remaining 9% replying *Maybe*. A total of 11 individual sessions were delivered, three received 100% excellent feedback and

D4.6 – Final report on dissemination and training

10 sessions has combined *Good* and *Excellent* rating of 82% or higher. One project received lower combined *Good* and *Excellent* rating of 64%.

Asked what the best part of the workshop was, respondents frequently mentioned a) the pace of the learning set by progressively more challenging projects b) visit to the computing facility c) Interaction between mentors and trainee. Similarly, we asked for the worst part of the workshop, respondents frequently mentioned a) project 2 - data clean up, though some respondents agreed that it was a necessary skill.

In total we are very satisfied with how the course went, the low number of attendants in the room was disappointing. Going forward, we will consider whether this course could be delivered online to lower the barrier of attendants; however, the setup of the course is not very compatible with online or virtual delivery as the quality of the teaching very much depended on the interaction between the participants as well as the interaction with the mentors.

Hands-on Introduction to HPC for Life Scientists (Joint BioExcel and PRACE)

The fifth BioExcel course “Hands-on Introduction to HPC for Life Scientists” was co-organised with PRACE and took place on 29th November 2017 to 1st December 2017 in Edinburgh, United Kingdom.

Course Overview

This course provided an introduction to High-Performance Computing (HPC) for researchers in the life sciences, using ARCHER as a platform for hands-on training exercises.

The course was co-organised and funded by BioExcel and PRACE (<http://www.prace-ri.eu/>), and delivered in collaboration with ARCHER – the UK national supercomputing service (<http://archer.ac.uk>).

High-performance computing (HPC) is a fundamental technology used to solve a wide range of scientific research problems. Many important challenges in science such as protein folding, the search for the Higgs boson, drug discovery, and the development of nuclear fusion all depend on simulations, models and analyses run on HPC facilities to make progress.

This course introduces HPC to life science researchers, focusing on the aspects that are most important for those new to this technology to understand. It will help participants judge how HPC can best benefit their research, and equip them to go on to successfully and efficiently make use of HPC facilities in future. The course covered basic concepts in HPC hardware, software, user environments, filesystems, and programming models. It also provided an opportunity to gain hands-on practical experience and assistance using an HPC system (ARCHER, the

D4.6 – Final report on dissemination and training

UK national supercomputing service) through examples drawn from the life sciences, such as biomolecular simulation.

Audience

This course follows on naturally from the BioExcel Summer School on Foundation skills for HPC in computational biomolecular research¹⁶. It is not necessary to have attended both courses.

Familiarity with basic Linux commands (at the level of being able to navigate a file system) was strongly recommended. Prospective applicants were referred to the following Linux 'cheat sheet' in case they were less familiar with Linux¹⁷. No programming skills or previous HPC experience is required.

Desktop computers running Windows were provided for the course, though participants were encouraged to bring their own laptop (running Windows, Linux, or macOS) as it could be useful to learn how to set up a connection to ARCHER (with assistance from course helpers if needed) and perform the hands-on practicals.

Learning outcomes

On completion of the course, we expect that participants will understand and be able to explain:

- Why HPC? – What are the drivers and motivation? Who uses it and why?
- The UK & EU HPC landscape – HPC facilities available to researchers
- HPC hardware – Building blocks and architectures
- Parallel computing – Programming models and implementations
- Using HPC systems
- Access
- Batch schedulers & resource allocation
- Running jobs
- Dealing with errors
- Compiling code
- Using libraries
- Performance
- The Future of HPC

Programme

Table 8. Programme for “Hands-on Introduction to HPC for Life Scientists (co-organised by PRACE and BioExcel), 29th November 2017 to 1st December 2017 in Edinburgh, United Kingdom.

Time	Topic	Speaker
------	-------	---------

¹⁶ <http://bioexcel.eu/events/bioexcel-summer-school/>

¹⁷ <http://www.archer.ac.uk/documentation/user-guide/linux.php#quickref>

D4.6 – Final report on dissemination and training

Day 1 – 29th November 2017		
10:00 - 11:00	Welcome, introduction, and course overview Introduction to BioExcel Review of HPC skills and competencies framework PRACE, ARCHER and EPCC training Familiarisation with fellow attendees	Vera Matser Arno Proeme David Henty
11:00 – 11:20	LECTURE - What is HPC?	Arno Proeme
11:20 - 11:30	PRACTICAL - Connecting to ARCHER	Arno Proeme
11:30 - 12:00	BREAK - Coffee & Tea	
12:00 - 13:00	PRACTICAL - Sequence Alignment	
13:00 - 14:00	BREAK - Lunch	
14:00 - 14:30	LECTURE - Parallel Computing Patterns	Arno Proeme
14:30 - 15:00	LECTURE - Measuring Parallel Performance	Neelofer Banglawala
15:00 - 15:30	PRACTICAL - Sequence Alignment (continued)	
15:30 - 16:00	BREAK - Coffee & Tea	
16:00 - 16:15	PRACTICAL - Sequence Alignment (continued)	
16:15 - 16:45	LECTURE - Building Blocks - Software (Operating System, Processes and Threads)	Michael Bareford
16:45 - 17:15	LECTURE - Building Blocks - Hardware (Processors & cores, Memory, Accelerators)	Juan Rodriguez Herrera
17:15 - 17:30	Review of the day	
17:30	End of day 1	
Day 2 – 30th November 2017		
09:30 – 09:45	Summary of day 1	
09:45 – 10:30	LECTURE - Parallel Models	Gordon Gibb
10:30 - 11:00	PRACTICAL - Fractal	
11:00 – 11:30	BREAK - Coffee & Tea	
11:30 - 12:00	PRACTICAL - Fractal (continued)	
12:00 - 12:30	LECTURE - HPC Architectures	Neelofer Banglawala

D4.6 – Final report on dissemination and training

12:30 - 13:00	LECTURE - Batch Systems & Parallel Application Launchers	Gordon Gibb
13:00 - 14:00	BREAK - Lunch	
14:00 - 15:00	PRACTICAL - Molecular Dynamics	
15:00 - 15:30	LECTURE - Compilers and Building Software	Michael Bareford
15:30 - 16:00	BREAK - Coffee & Tea	
16:00 - 16:30	PRACTICAL - Molecular Dynamics (continued)	
16:30 - 17:00	LECTURE - Parallel libraries	Juan Rodriguez Herrera
17:00 - 17:15	Review of the day	
17:15	End of day 2	
19:00	Dinner at Amber restaurant	
Day 3 – 1st December 2017		
09:30 - 09:45	Summary of day 2	
09:45 - 10:15	LECTURE - Pipelines and workflows	Arno Proeme
10:15 - 11:00	PRACTICAL - Molecular Dynamics (continued)	
11:00 - 11:30	LECTURE - The UK & EU HPC Landscape	Arno Proeme
11:30 - 12:00	BREAK - Coffee & Tea	
12:00 - 12:30	LECTURE - The Future of HPC	Arno Proeme
12:30 - 13:00	LECTURE - "Where next?" and things to remember	Arno Proeme
13:00 - 14:00	BREAK - Lunch	
14:00 - 15:00	Individual consultations, course review and feedback / competency survey	Vera Matser
15:00	End of course	

Conclusion and course feedback

The Hands-on Introduction to HPC for Life Scientists was jointly organised by BioExcel and PRACE. The registration for this course ran through the PRACE registration system on a first come, first served basis with the maximum capacity of the course set at 25 participants. The attendance rate for the course was 72%, and the gender balance during the workshop 39% male and 61% female. The course was hampered by a high no-show and cancellation rate, a total of 18 people

D4.6 – Final report on dissemination and training

either cancelled (15) or did not show up on the day (3). For a number of the earlier cancellations we were able to contact someone on the waiting list, the closer to the event the harder this becomes. Counting the cancellations, the application gender balance was 42% male and 58% female.

The course feedback was collected through the standard PRACE feedback form, complemented by a few additional questions to bring it in line with the BioExcel feedback form. The response rate for the feedback survey was 94%, when asked for an overall rating for the course, 94% of respondents rated the course as *Good* or *Excellent* (65% *Excellent*). 100% of the respondents rated the overall course organisation as *Good* or *Excellent*. When asked about the balance between theoretical and practical content 88% replied *About right*, the remaining 12% answered that the course was too theoretical. 94% of respondents would recommend the course to their colleagues. Using the following rating options don't know < disagree completely < disagree < no strong feelings < agree < agree completely, respondents were presented with the statement "the lectures were clearly presented and comprehensible" and "the pace of teaching was right". On clarity, 18% responded with *agree completely*, 65% *agreed* and 18% responded with *no strong feelings*. Similarly, for the pace of teaching 29% responded with *agree completely*, 53% *agreed*, and 18% responded with *no strong feelings*.

Asked what participants liked most about the course, respondents frequently mentioned a) practical exercises b) the teaching staff was very helpful. Similarly, we asked what they disliked about the course and respondents frequently mentioned a) some lectures/practicals still too technical/computationally focused.

This is the first time we have run this course specifically for life scientists, the course was clearly successful, but we think there is room for improvement; in the next iteration of the course we will continue to work on making the life scientist the focus of the sessions and exercises.

BioExcel Summer School on Biomolecular Simulations 2018

The BioExcel Summer School 2018 on Biomolecular Simulations took place on 17th - 22nd June 2018 in Sardegna Ricerche, Pula, Italy. The summer school organising committee consisted of BioExcel members Vera Matser (EMBL-EBI) and Alexandre Bonvin (Utrecht University), as well as the local hosts Attilio Vittorio Vargiu and Giuliano Mallocci from the University of Cagliari.

Course Overview

The summer school provided detailed training on the BioExcel codes (GROMACS, HADDOCK, CPMD, PMX) as well as methods and best practice. The summer school included lectures and hands-on sessions on the following topics:

- Molecular Dynamics simulations
- Biomolecular Docking

D4.6 – Final report on dissemination and training

- QM/MM
- Free energy calculations
- Advanced sampling methods (metadynamics)

During the hands-on computer practicals participants worked on a use case integrating the various topics above making use, among others, of the BioExcel flagship software (GROMACS, HADDOCK, PMX, CPMD).

Audience

The summer school was intended for researchers (primarily PhD and post-docs) using or planning to use biomolecular modeling and simulation in their everyday research. Familiarity with Linux and some basic knowledge of molecular modelling software was a requirement. The summer school was an open application with selection. The selection was carried out by the summer school organising committee.

Learning outcomes

After this course you should be able to:

- Use a range of biomolecular modeling and simulation software (GROMACS, HADDOCK, PMX, CPMD)
- Discuss current trends and challenges in biomolecular simulation

Programme

Table 9. Programme for the BioExcel Summer School 2018 on Biomolecular Simulations on 17-22 June 2018 in Sardegna Ricerche, Pula, Italy.

Time	Topic	Speaker
Day 0 - Sunday 17th June 2018		
19:30	Registration	
19:45	Dinner @ Hotel Flamingo	
Day 1 - Monday 18th June 2018		
08:30 - 09:00	Bus to the venue	
09:00 - 09:30	Welcome & Introduction to BioExcel	Vera Matser & Alexandre Bonvin
09:30 - 10:30	Lecture - Introduction to molecular dynamics and its workflows	Mark Abraham
10:30 - 11:00	Coffee & tea break	

D4.6 – Final report on dissemination and training

11:00 - 12:30	Lecture - Introduction to GROMACS and some of its algorithms	Mark Abraham, Paul Bauer & Viveca Lindahl
12:30 - 13:30	Lunch	
13:30 - 14:30	Flashtalks	
14:30 - 15:30	Tutorial - First steps in using GROMACS	Mark Abraham & Paul Bauer
15:30 - 16:00	Coffee & tea break	
16:00 - 17:30	Tutorial - Using AWH enhanced sampling in GROMACS	Viveca Lindahl
17:30 - 18:00	Bus back to the hotel	
19:30 - 21:30	Dinner @ Hotel Flamingo	
Day 2 - Tuesday 19th June 2018		
08:30 - 09:00	Bus to the venue	
09:00 - 10:30	Lecture - Modeling molecular recognition through enhanced-sampling methods	Attilio Vargiu & Andrea Basciu
10:30 - 11:00	Coffee & tea break	
11:00 - 12:30	Lecture - Integrative modelling of biomolecular complexes with HADDOCK	Alexandre Bonvin
12:30 - 13:30	Lunch	
13:30 - 14:30	Flashtalks	
14:30 - 15:30	Tutorial - HADDOCK	Alexandre Bonvin & Adrien Melquiond
15:30 - 16:00	Coffee & tea break	
16:00 - 17:30	Tutorial - HADDOCK (continued)	Alexandre Bonvin & Adrien Melquiond
17:30 - 18:00	Bus back to the hotel	
19:30 - 21:30	Dinner @ Hotel Flamingo	
Day 3 - Wednesday 20th June 2018		
08:30 - 09:00	Bus to the venue	
09:00 - 10:30	Lecture - QM/MM interfaces in the CPMD code	Emiliano Ippoliti
10:30 - 11:00	Coffee & tea break	

D4.6 – Final report on dissemination and training

11:00 - 12:30	Tutorial - CPMD	Emiliano Ippoliti & Viacheslav Bolnykh
12:30 - 13:30	Lunch	
13:30 - 14:30	Tutorial – CPMD (continued)	Emiliano Ippoliti & Viacheslav Bolnykh
14:30 - 16:00	Poster session & Coffee and tea break	
16:00 - 20:00	Travel to Nora - Excursion to Nora	
20:00 - 22:00	Social dinner in Pula	
22:00 - 22:30	Travel back to the hotel	
Day 4 - Thursday 21st June 2018		
08:30 - 09:00	Bus to the venue	
09:00 - 10:30	Lecture - Alchemical free energy calculations using PMX	Bert de Groot
10:30 - 11:00	Coffee & tea break	
11:00 - 12:30	Tutorial - PMX	Bert de Groot & Vytautas Gapsys
12:30 - 13:30	Lunch	
13:30 - 14:30	Tutorial – PMX (continued)	Bert de Groot & Vytautas Gapsys
14:30 - 15:30	Tutorial - Metadynamics / HADDOCK	Andrea Basciu & Adrien Melquiond
15:30 - 16:00	Coffee & tea break	
16:00 - 17:30	Tutorial - Metadynamics / HADDOCK	Andrea Basciu & Adrien Melquiond
17:30 - 18:00	Bus back to the hotel	
19:30 - 21:30	Dinner @ Hotel Flamingo	
Day 5 - Friday 22nd June 2018		
08:30 - 09:00	Bus to the venue	
09:00 - 10:00	Research highlights - short talks from participants	Lara Callea, Joan Francesc Gilabert Navarro & Ksenia Korshunova
10:00 - 10:30	Talk - Dynamical structure-function relationship in bacterial transporters	Paolo Ruggerone

D4.6 – Final report on dissemination and training

10:30 - 11:00	Coffee & tea break	
11:00 - 12:15	Round table & get answers session	
12:15 - 12:30	Final remarks and feedback	
12:30 - 13:30	Lunch	
14:30	Travel to the airport	

Conclusion and Course Feedback

The Summer School 2018 was a very successful event, a total of 66 people applied (41% female applicants, 59% male applicants) for 30 places. We received a total of 15 requests for travel grants. All applicants were ranked by the organising committee, the criteria were suitability for the course and the ability to apply the knowledge immediately. The travel grants were awarded within the top 30 applicants, we excluded 2 applicants based on geography and career stage, giving preference to applicants with a higher ranking. The top 30 applicants received an invitation, all applicants accepted their place and registered for the course. The gender balance for the participants on the course was 57% female and 43% male. We had 100% attendance on this course (places filled/maximum places available). We expect a large factor of this to be due to the fact that a registration fee was charged to part cover the hotel accommodation costs.

All successful participants were invited to bring a poster and present a 1-minute flash talk. Four participants were invited to present a short talk on the last day of the summer school.

The standardised BioExcel feedback survey was used to ask participants for their feedback on the event. The feedback percentage was 90%. The participant feedback was very positive, 100% of the respondents rated the overall events as *Good* or *Excellent* (65% rated *Excellent*), the overall workshop organisation was also rated *Good* or *Excellent* by 100% of the respondents (80 rated *Excellent*) and 100% would recommend the summer school to colleagues. A total of 16 individual sessions were delivered; five sessions had 100% rating in *Good* and *Excellent*, only for two sessions did the combined *Good* and *Excellent* rating fall below 70% (55% and 65%). For both these sessions we have received sufficient specific feedback from the participants to isolate the reason for the lower scores and feel confident that we will be able to improve these sessions. The balance between theoretical and practical content was considered *About right* by 95% of the respondents.

Asked what the best part of the workshop was, respondents frequently mentioned a) Amount of time for direct interaction with the speakers (note that all speakers attended the full summer school and were housed in the same hotel complex as the participants. All dinners were taken together as a group), b) networking with other participants, c) the social excursion to Nora and the workshop dinner (100% excellent, with rave reviews for the guide at the Nora excursion), d) coverage of a broad set of techniques, e) individual sessions such as HADDOCK received very

D4.6 – Final report on dissemination and training

positive mentions. Similarly, we asked for the worst part of the workshop, respondents frequently mentioned the following a) nothing rated as worst, b) organisation of the poster session could be improved, c) quality of the wifi in the hotel, d) improvement suggestions for the sessions that received the lowest feedback.

Travel grants

BioExcel offers a limited number of fixed amount travel grant for each of the BioExcel Training events. Full terms and conditions¹⁸ are made available to prospective applicants on the events page of each of the BioExcel Training courses. When selecting the travel grant beneficiaries, we generally take into account gender, geographic spread, and how much impact the course would have on the applicant's research.

Table 10 provides an overview of the people who have been awarded travel grants for the BioExcel training courses.

*Table 10. BioExcel training course travel grant beneficiaries. *Due to unforeseen circumstances this person needed to cancel their participation in the training event and subsequently the grant was withdrawn.*

Last Name	First Name	Company Name
20-21st October 2016 - BioExcel: workflow training for computational biomolecular research		
Camacho Cano	Esther	Centro de Biología Molecular Severo Ochoa
Flores	Antonio	Universidad de Málaga
<i>Hernandes Coutinho*</i>	<i>Felipe *</i>	<i>Universidade Federal do Rio de Janeiro *</i>
Matsoukas	Minos	University of Patras
Gurinova	Jana	University of Vienna
10-13th April 2016 - PRACE/BioExcel Spring School - HPC for Life Science		
Kravchuk	Vladyslav	Taras Shevchenko National University
Llorach Pares	Laura	University of Barcelona - Mind the byte
Rutkowska	Ewelina	University of Warsaw
3-7th July 2017 - BioExcel Foundation Skills for HPC in Computational Biomolecular Research		
Thakker	Alpesh	University of Birmingham
Camisasca	Gaia	Stockholm University
Tsengenes	Alexandros	Biomedical Research Foundation of Academy of Athens (BRFAA)
Aquino	Genevieve Mae	Philippine Genome Centre - Agriculture, University of the Philippines

¹⁸ <http://bioexcel.eu/wp-content/uploads/2017/01/BioExcelTravelGrant.pdf>

D4.6 – Final report on dissemination and training

18-22 nd July 2018 - BioExcel Summer School on Biomolecular Simulations		
Melarkode Vattekk	Akhia	INSERM UMR_S1134
Prates	Nubia	Federal University of Minas Gerais
Shehnaz Begum	Saheen	Tezpur University
Barrera-Vilarmau	Susana	Institute of Advanced Chemistry of Catalonia (I.Q.A.C.), Spanish Council for Scientific Research (C.S.I.C.)
Elisée	Eddy	Institut de Chimie des Substances Naturelles

The beneficiaries of the BioExcel travel grant are asked to write a blog post for us, either about their experiences at the course, their research interests (in layman's terms) or a relevant scientific topic. The blogs are published on the BioExcel website¹⁹ and promoted via twitter. In Figure 8 we show one example of a travel grant blog published on our website, written by Eddy Elisée from Institut de Chimie des Substances Naturelles about the BioExcel Summer School on Biomolecular Simulations.

Eddy Elise

Institut de Chimie des Substances Naturelles



More than a workshop, BioExcel Summer School is a human experience. It allows to meet people working in your field from all over the world (PhD students, Post- doctoral fellows, researchers, team leaders) in a friendly atmosphere. It is the kind of event where you learn, teach, share and exchange the scientific (or not) knowledge.

This summer, BioExcel Summer School took place in Pula, in the wonderful island of Sardinia (Italy). We spent our time between the Research Centre and the hotel, near the beach. The program was a mix of expert lectures, well- explained tutorials and student presentations where all questions found an answer.

I really enjoyed the whole summer school since I learnt a lot of things and tips about how to use new bioinformatic tools and how to improve those I frequently use during my PhD. As far as I am concerned, this latter point is very important and exciting since this event gives you the opportunity to directly approach developers or creators of well-known software. You can easily ask questions, state your issues or more interestingly, discuss what you developed using from the current version of the software.



Above all, I met exceptional people coming from Spain, Switzerland, England, Italy, Russia, Armenia, Czech Republic and Netherlands. There was no barrier between organizers and participants: everyone talked with everyone, at lunches, social dinners, during a walk in the ancient city of Nora, or while chilling on the beach. I really recommend people to attend one of the amazing and well-organised BioExcel events in order to find new international friends, new collaborations, or, in general, to enrich your network which is our best ally in Science since together we are stronger than alone.

Let's share science!

Figure 8. Travel grant blog published on our website²⁰, written by Eddy Elisée from Institut de Chimie des Substances Naturelles about the BioExcel Summer School on Biomolecular Simulations.

¹⁹ <https://bioexcel.eu/category/testimonials/>

²⁰ <https://bioexcel.eu/hear-from-our-travel-grant-beneficiaries-2/>

Impact Assessment

To ensure the quality of our training programme BioExcel routinely collects course feedback for all BioExcel training activities. The responses provide us with valuable information on how to improve our events and the direction of the training programme. However, these responses are generally collected on the last day of the event and therefore do not provide us with a lot of information about the long-term impact of the BioExcel training programme and tend to focus on immediate concerns (e.g. a problem with a tutorial, a session was too short or too difficult). To assess the impact of our training programme activities, we want to know if attending our training courses has had a positive impact on their research/career and whether or not they are actively using the tools that were introduced to them during the training event.

To assess the long-term impact, BioExcel sends two long-term feedback surveys to all delegates who have attended BioExcel training courses, the first one is sent approximately 6-12 months after the training event, the second survey is sent after 2 years. A similar survey has been used by EMBL-EBI since 2010 and has been adopted by ELIXIR as well. The template for the 6-12 month survey was included in D4.5 - Training report and updated plan²¹. The surveys collect data on whether participants are using the tools they were introduced to during the training course, whether they have established collaboration, and whether they have taught others. In addition, we ask what impact the course has had on their research and the research of other, as well as what the best and worst parts of the course was.

All course data is combined and analysed collectively. At this point in time we do not specifically analyse data at a course-by-course basis, though this may be something that we can consider in the future once more data is available.

Response rate long-term feedback surveys

The response rate for our long-term feedback survey is one of the WP4 KPIs (WP4_KPI05_B). The average response rate for our 6-12 months survey is 26% which is very good, our stretch target was 25%. There is a lot of variety in the response rate between the courses; for event 2 the response rate is below the reachable KPI target of 15%, and for event 4 is it very high at 55%. Only two of our events are far enough in the past for us to have sent out the 2-year feedback from. The average response rate for the 2-year is 17.5% which both events have very similar response rates.

Table 11. Response rate for the long-term feedback surveys broken down per course.

Course	Number of students received survey	Response rate for 6-12 months	Response rate for 2 years
--------	------------------------------------	-------------------------------	---------------------------

²¹ <https://doi.org/10.5281/zenodo.574620>

D4.6 – Final report on dissemination and training

Event 1	6	33%	17%
Event 2	17	6%	18%
Event 3	53	19%	-
Event 4	11	55%	-
Event 5	18	17%	-
Event 6	-	-	-
Total	105	26%	17.5%

Long-term feedback survey 6 - 12 month

There are a number of elements that we look at when we try to determine the impact of our training activities; we are interested in whether participants have passed on their new skills, whether they have made any new collaborations, whether they would and have recommended the course to others and what impact they feel the course has had on their research and the research of others.

We encourage our participants to share their skills with colleagues or students. The vast majority of our participants (84%) teach others the skills they have learnt. When asked “how many others have you taught”, 79% have taught between 1-5 others, and 5% have taught 20+ (Figure 9). We consider this to be a really positive impact as it shows that the number of people benefitting from our training is significantly higher than the number of people in the classroom.

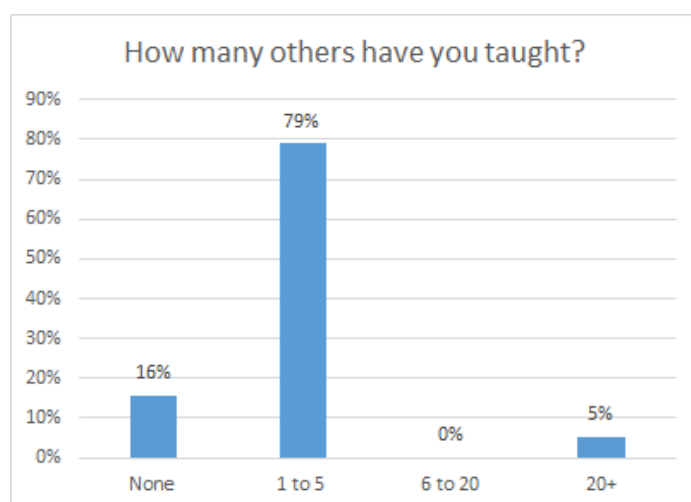


Figure 9. Response from the 6-12 months long-term feedback survey asking, “How many others have you taught?”.

Networking is an essential part of our training courses, when designing our courses, we make sure that there is sufficient time for coffee breaks and where possible joint course dinners are organised in the evening. Where possible the

D4.6 – Final report on dissemination and training

trainers are always encouraged to join the social activities around the course. Helping to expand participants network is a positive outcome by itself, and highly valued by the participants (see figure 12); sometimes expanding the networks leads to collaborations, either between participants or with BioExcel researchers (see Figure 10). When asked whether participants have established collaborations with participants or BioExcel researchers 21% of the respondents replied *Yes*. As a Centre of Excellence, we consider facilitating these collaborations as a core part of our mission and we are very pleased to see this substantiated in our long-term feedback results.

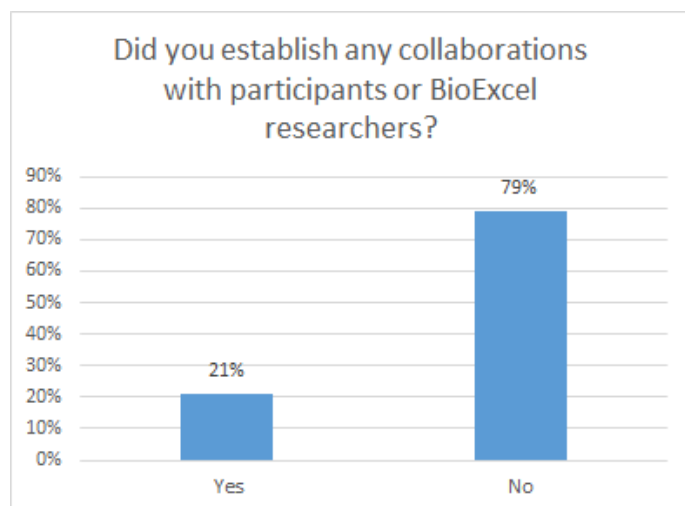


Figure 10. Response from the 6-12 months long-term feedback survey asking "Did you establish any collaborations with participants or BioExcel researchers?"

We ask our former participants if they would recommend our course, or if they have already recommended it. In our post-course feedback survey we get very high scores on this question, but it is important to understand whether after reflection our former participants would recommend our courses. We are very pleased to see that the vast majority (90%) would recommend the course they have attended and many already have recommended our courses to others (58%).

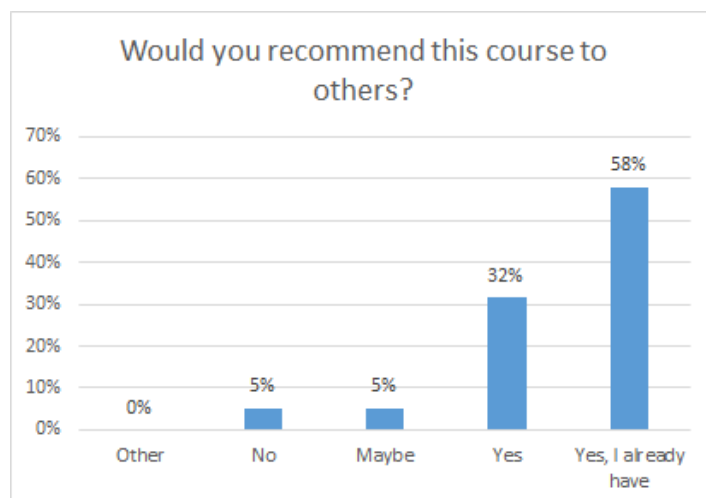


Figure 11. Response from the 6-12 months long-term feedback survey asking "Would you recommend this course to others?"

D4.6 – Final report on dissemination and training

In addition to a number of pre-formulated questions, we also have an open-ended question in our survey “What impact has the course had on your research and the research of others?”. To analyse the impact statements we try and find common themes across the participants statements. We score each statement across the commonly identified themes.

See an example of one of the impact statements we have received below:

“I wouldn't have conducted my research without it”

This statement has been scored against the “Essential to research/actively using skills” theme.

If a new theme is identified, we re-analyse past statements. Note that participants impact statements often contains multiple commonly identified themes. In total 8 common impact themes were identified (Figure 12). The most commonly reported theme is “Essential to research/actively using skills” with 10 counts, followed by “Gained knowledge/skills” with 8 counts; these themes are related but the second one is weaker and doesn't necessarily mean active use of the skills. These impact themes are core to training activities and we are encouraged to see them front and center in the analysis. Another adjacent theme is “Encouraged further learning/overcame fear”, here participants indicate that they used the training course as a starting point to learn more, we expect this to be more common for introductory subjects. On occasion, participants specifically state that the course helped them to overcome their fear of a subject, and they are now learning more about it or are actively using the skill. The counts for “Expanding my network” and “I taught others” are underrepresented since there is a separate question related to this in the survey. Impact themes such as “Influenced a career decision”, “Better able to communicate with experts” and “Published a paper” are mentioned less frequently.

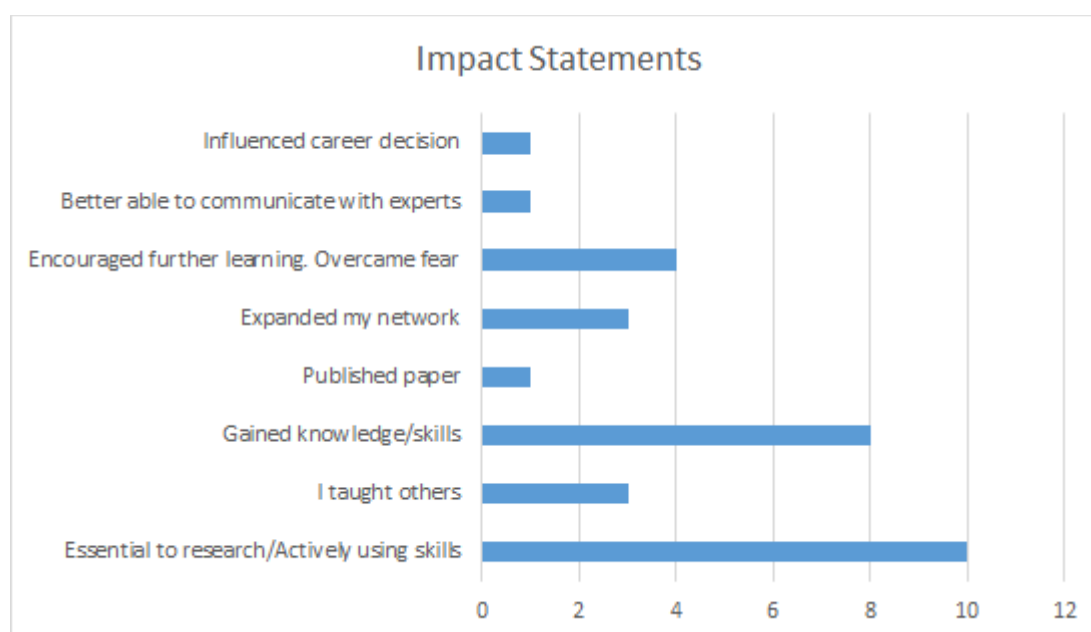


Figure 12. Analysis of the impact statements collected in the 6-12 months long-term feedback survey.

Long-term feedback survey 2 years

Participants who have attended our 2016 courses have been sent the 2-year long-term feedback survey. A detailed analysis of the 2-year survey data does not make a great deal of sense as we only have 4 responses in total. The questions asked are similar to the 6-12 months survey, but in addition we ask participants if they were working in the field before the course and after the course. This will allow us to track how people might be exiting the field after 2 years. The current results have one person (25%) exiting the field. We also explicitly ask whether participants have written a dissertation/thesis following their participation in the course, and whether they have published a paper following their participation in the course; the current respondents have replied *No* to both questions.

At the 2-year mark, we are additionally interested in the career development of our attendees; we therefore ask the following questions “Did you receive/apply for a grant after attending this course?” (25% Yes, 75% No) and “Did this course lead to a change of position or other career development?” (100% No). For the impact statements we have counted 1 time “Essential to research / Actively using skills” and 2 times “Gained knowledge/skills”.

Impact assessment conclusion

It is hard to quantify the long-term impact of training programmes, we believe that our long-term feedback survey helps us to determine the impact. We will gain a more complete picture once all of our courses have passed the 2-year mark. The impact assessment activities will continue to run under BioExcel-2 allowing us to complete the assessment. In future we will consider having more than one person score the impact statements across common themes to provide a more unbiased view. We have in the past considered whether to include predefined impact statements in the survey, these would be based on the historic responses we have collected. However, we are afraid this could either lead people in a certain direction or that respondents would not list impact that is not part of the predefined statements.

Sustainability

In this section we present the work carried out to make the training activities of BioExcel more sustainable, this includes making our training material and resources more findable through the BioExcel Knowledge Resource Center, but also working with WP5 to analyse the potential of training as a sustainable product.

Knowledge Resource Center

The Knowledge Resource Centre is a database of computational biomolecular training courses aggregated from BioExcel partners and external providers. Through the previous years, a total of 156 courses were collated, out of which 131

D4.6 – Final report on dissemination and training

courses were taught online and 25 courses were taught face-to-face. After a manual curation, 19 courses were found to have broken web links and were not available online.

Initial Mapping and Gap analysis

The aim of the mapping exercise is not to find all relevant training courses but to identify where the gaps are. The available training resources were mapped to their relevant competencies based on the course description as per table 12. As the information is solely based on the course description, it is important to keep in mind that this process does not take into account the quality or frequency of the resources.

Table 12. Description of competencies and mapped courses

Competency	Courses mapped
1.1 Function effectively in teams to accomplish a common goal.	1
1.2 Comprehend and comply with professional, ethical, legal, security, and social issues and responsibilities, and uphold these in the workplace as appropriate.	1
1.3. Communicate effectively with a range of audiences, technical and nonprofessional.	1
1.4. Engage in continuing professional development	1
1.5. Identification and understanding of user needs.	5
2.1 Apply computing expertise appropriate to the discipline and level of expertise (software as well as hardware).	20
2.2 Apply expertise in formal & natural sciences appropriate to the discipline.	29
2.3 Detailed comprehension of the scientific process and an ability to apply it (e.g. experimental design; inclusion of controls; QA, QC, processing, visualisation and interpretation of results).	16
2.4 Comprehension of, and compliance with, licensing policy.	3
2.5 Use a computer-based system, process, component, or program to meet desired needs in a biomolecular context.	11

D4.6 – Final report on dissemination and training

2.6 Evaluate the ability of a computer-based system, process, component, or program to meet desired needs in a biomolecular context.	11
2.7 Comprehension of the local and global impact of high-performance computing (HPC) and high-throughput computing (HTC) on individuals, organizations, and society.	4
2.8 Comprehension of how data-driven science, data analysis and computational modelling can be combined to generate and test hypotheses (e.g. machine learning, data mining, pattern recognition).	15
2.9 Identify and compile appropriate data sets to address specific research questions	15
2.10 Comprehension of, and compliance with, best practice in data management / organization / archiving and storage	11
2.11 Comprehension of, and compliance with, best practice in distributed data management and data management planning	20
2.12 Search for, read and assess literature / manual	1
2.13 Presenting your results to community (writing papers, conference presentation, YouTube)	3
3.1 Comprehension of, and compliance with, good programming practice (as promoted, for example, by www.software-carpentry.org).	5
3.2 Analyze a problem and identify and define the computing requirements appropriate to its solution (e.g., define algorithmic time and space complexities and hardware resources required to solve a problem).	10
3.3 Apply knowledge of the operating system.	6
3.4 Write/adapt computer programs (software development) for biomolecular simulations.	10
3.5 Write his/her own scripts to perform tasks in context of biomolecular research.	13
3.6 Install biomolecular simulation software on his/her computer.	5

D4.6 – Final report on dissemination and training

3.7 Deploy and test non-commercial software, including software that is built collaboratively and on a volunteer basis.	1
3.8 Apply knowledge of systems monitoring (e.g. queue monitoring, systems availability and optimisation, storage used; scheduling maintenance at appropriate times and communicating this to users).	8
4.1 Assess computational workflow systems and their potential benefits.	9
4.2 Apply knowledge of batch system	11
4.3 Write computer programs that can run on a parallel computer	7
4.4 Assess advantages and limitations for deploying, executing and optimising computations in a cloud/grid/HPC environment	14
4.5 Apply knowledge of performance profiling to measure suitability of computing platforms	15

34 new resources were added by conducting an online search using a combination of the search terms such as computational science, structural biology, simulations and drug discovery (see Figure 13). Moving forward, including a targeted search of resources to minimize gaps would be conducted.

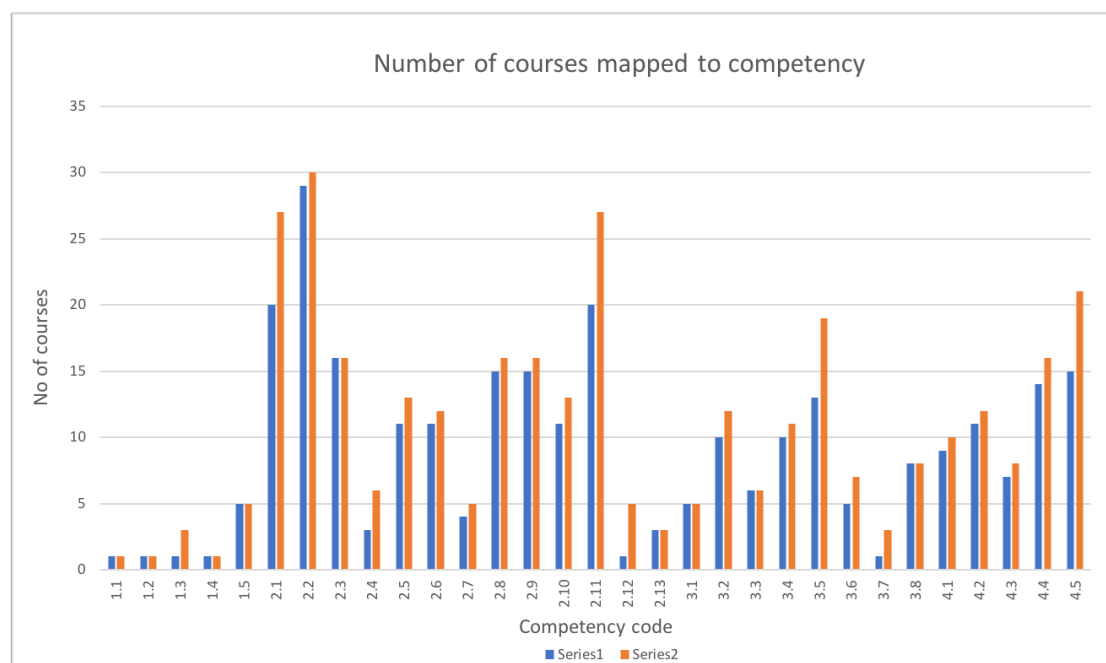
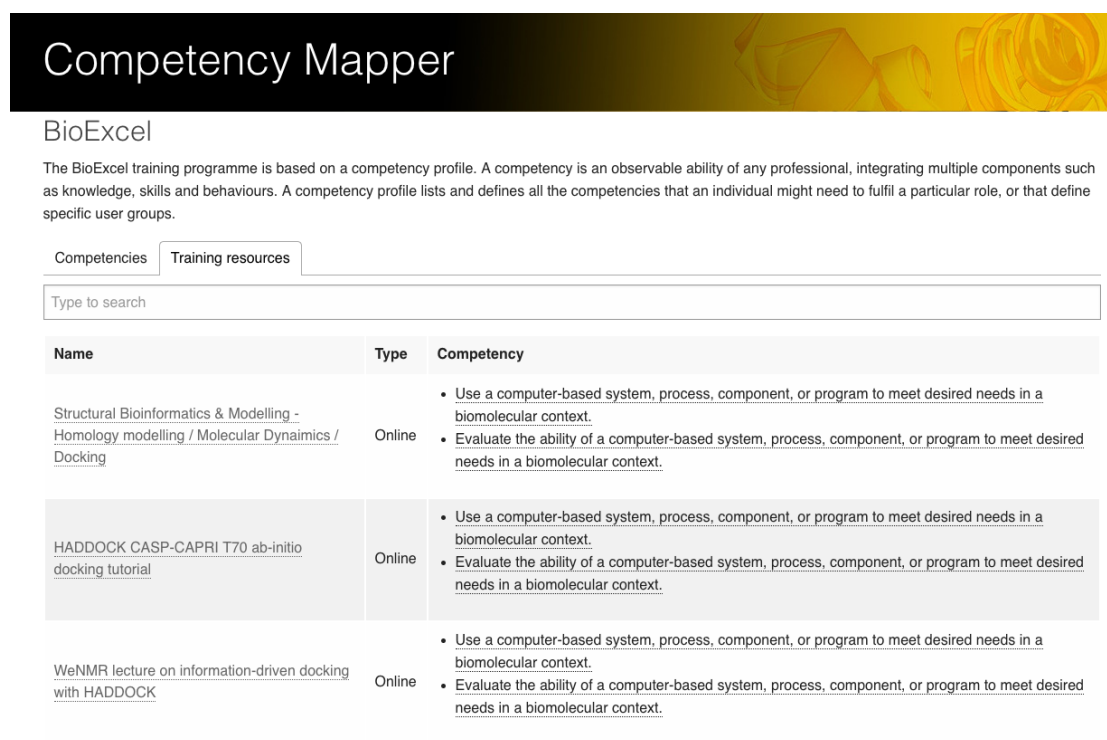


Figure 13: Updated list of training resources. Series 1 represents graph with original list of courses, Series 2 represents graph with new courses added.

Competency Mapper

The BioExcel partner EMBL-EBI is involved in a number of initiatives that use competency profiles; a competency mapper was designed by EMBL-EBI with BioExcel as a use case²². The competency mapper was designed, both to give a permanent home to these competency frameworks, regardless of the project/initiative end date, as well as provide an easy and consistent method to conduct and visualise the mapping of training resources to competencies. The competency mapper was introduced to map courses to the framework including knowledge, skills and attributes. Below are the listed steps in order to add a training resource to the site.

- 1) The homepage displays a list of projects hosted by the site. Select 'BioExcel' and 'manage training resources' to add a new course.



BioExcel

The BioExcel training programme is based on a competency profile. A competency is an observable ability of any professional, integrating multiple components such as knowledge, skills and behaviours. A competency profile lists and defines all the competencies that an individual might need to fulfil a particular role, or that define specific user groups.

Competencies Training resources

Type to search

Name	Type	Competency
Structural Bioinformatics & Modelling - Homology modelling / Molecular Dynamics / Docking	Online	<ul style="list-style-type: none">• Use a computer-based system, process, component, or program to meet desired needs in a biomolecular context.• Evaluate the ability of a computer-based system, process, component, or program to meet desired needs in a biomolecular context.
HADDOCK CASP-CAPRI T70 ab-initio docking tutorial	Online	<ul style="list-style-type: none">• Use a computer-based system, process, component, or program to meet desired needs in a biomolecular context.• Evaluate the ability of a computer-based system, process, component, or program to meet desired needs in a biomolecular context.
WeNMR lecture on information-driven docking with HADDOCK	Online	<ul style="list-style-type: none">• Use a computer-based system, process, component, or program to meet desired needs in a biomolecular context.• Evaluate the ability of a computer-based system, process, component, or program to meet desired needs in a biomolecular context.

²² <https://ebiwd.github.io/competency-mapper/#/>

D4.6 – Final report on dissemination and training

Create Training Resources

Create new competency

Title
Title

Start Date dd/mm/yyyy **End Date** dd/mm/yyyy

Event type
Online

Description

X Copy Paste Undo Redo Bold Italic Text color Background color Link Unlink Source

B I S Ix | | | | | Styles - | Format - | ?

Location
Location

3) Select 'manage competency profile' to allocate relevant competencies

Manage competency profile for BIOEXCEL

1	Generic Competencies
	<input type="checkbox"/> Function effectively in teams to accomplish a common goal
	Knowledge
	<input type="checkbox"/> Aware of cultural differences.
	<input type="checkbox"/> Understanding of the context of the persons in the team
	<input type="checkbox"/> Knowledge of the domain
	Skill
1.1	<input type="checkbox"/> Communication
	<input type="checkbox"/> Conflict management
	<input type="checkbox"/> Time management
	Attitude
	<input type="checkbox"/> Invites two-way communication; actively listens/pays attention; able to excite participation and commitment from others
	<input type="checkbox"/> Delivers on his/her actions and inspires this behaviour in others
	<input type="checkbox"/> Informs others of relevant information appropriately and on time

4) The course is displayed on the site with its tagged competencies

D4.6 – Final report on dissemination and training

BioExcel

The BioExcel training programme is based on a competency profile. A competency is an observable ability of any professional, integrating multiple components such as knowledge, skills and behaviours. A competency profile lists and defines all the competencies that an individual might need to fulfil a particular role, or that define specific user groups.

Name	Type	Competency
Structural Bioinformatics & Modelling - Homology modelling / Molecular Dynamics / Docking	Online	<ul style="list-style-type: none">Use a computer-based system, process, component, or program to meet desired needs in a biomolecular context.Evaluate the ability of a computer-based system, process, component, or program to meet desired needs in a biomolecular context.
HADDOCK CASP-CAPRI T70 ab-initio docking tutorial	Online	<ul style="list-style-type: none">Use a computer-based system, process, component, or program to meet desired needs in a biomolecular context.Evaluate the ability of a computer-based system, process, component, or program to meet desired needs in a biomolecular context.

Training as a sustainable product

Training courses have been identified as potential product to build a sustainability strategy for BioExcel. Training is an activity that users are prepared to pay for and are to some extent reusable.

The initial thoughts around the training sustainability product have been expanded into a virtual training programme that will be piloted in BioExcel-2. In this context, virtual training means providing training with direct interaction between trainer and participants (and between participants, to some extent) facilitated by software such as GoToTraining and access to shared resources (likely via container), including compute resources. Expanding our training programme to include virtual training will allow us to expand our reach. Currently, though we do not charge registration fees, a multi-day course can still be prohibitively expensive for attendees when considering travel and accommodation costs. In addition, the trainer/support staff time is a limited commodity and virtual training would be less demanding. In the long-term, we also hope for selected courses to be able to support participants over a longer period to facilitate the integration of course knowledge into their daily research. It is important to note that we do not anticipate virtual training to replace face-to-face training, this will be an additional training modality.

The complexity of the pilot sessions will be built up over time. We anticipate starting with a one module pilot session that does not require additional compute requirements but can be run on a user's own laptop or PC, building to a multi-day event where additional computing infrastructure will be needed e.g. to visualise molecules using PyMOL. We anticipate that the course will use a container that has the course software installed (including software or links needed to deliver the course, such as GoToTraining), and any needed course material (e.g. the pre-recorded lectures, and tutorial files). Training in a virtual setting will be an adjustment for the trainers, we anticipate the initial pilot sessions to be relatively computationally undemanding to refine the virtual human-interaction.

Proposed structure for a virtual training module:

Live session 0: technical support session. We anticipate providing a checklist allowing users to test whether their system has been installed and configured correctly. If users encounter problems, support is available during this session to help them setup the system.

Recorded lectures: users can watch the recorded sessions in their own time

Live session 1: discussion of lectures by the main course scientific trainer, questions and answer session and introduction of the tutorial.

Online tutorial: users will have access to the tutorial and work through it in their own time

Live session 2: at specific times live support will be available for the tutorial. If users have not anticipated any problem they do not have to attend.

Note: Components in brackets can be repeated to chain multiple modules into a larger course.

Live session 3: final wrap up session to discussion and remaining issues, discuss challenges and limitations of the methods.

The activities will be aligned with the BioExcel sustainability planning to ascertain the commercial viability of this type of virtual training, (see D5.4 Governance and Business Plan²³). Note that we do not anticipate all virtual training pilot sessions to be developed into commercial products.

5. KPI

The key performance indicators were revised, and targets realigned at the beginning of the second reporting period. We have adjusted our reach and stretch targets, especially for outreach/dissemination, to make them more challenging. KPIs are shown in Annex 1.

6. Conclusion & future work

In the second reporting period we have continued to build on the BioExcel brand established during the first reporting period. We have run a very active dissemination and training strategy with contributions to over 40 events. We have a successful training programme that has brought recognition to BioExcel as a leader in the field of competency-based training. We are using the ELIXIR

²³ D5.4 – will be made available at <https://bioexcel.eu/research/deliverables/>

D4.6 – Final report on dissemination and training

methodology to analyse the impact of our training programme. We plan to continue our dissemination and training activities during BioExcel-2 and will continue to follow up the long-term impact assessment of the BioExcel-1 training courses during BioExcel-2. Plans are underway to deliver another iteration of the Spring School with PRACE, and a BioExcel Summer School in 2019; in addition, we will be piloting our virtual training programme. A major large conference is planned for 2020 as a major venue for the wider computational biomolecular community.

We are planning to continue building on our already established collaborative agreements and cement new agreements where appropriate. We look forward to working with the newly established Focus CoE on training and outreach. In the next phase of the funding project, we have a specific task on branding and development of quality mark to solidify BioExcel as a quality mark for excellence in computational biomolecular research.

Annex 1. KPIs

Table 13: KPI overview for WP4, the colour code indicates whether the KPI failed (pink), met the reachable target (light green) or met the stretch target (dark green)

ID	KPI	Justification	Reachable target	Stretch target	PM21	PM24	PM27	PM30	PM33	PM36
WP4_KPI02	Number of CoE training events	Training events need to cover the main areas of expertise in the centre	2 per year, 6 in total	8 in total	4	n/a, see PM21	5	n/a, see PM27	6	n/a, see PM33
WP4_KPI03	Number of people who attended BioExcel courses	Courses need to cover topics of interest for the users	70% of maximum number of places	90% of maximum places	44%	n/a, see PM21	64%	n/a, see PM27	100%	n/a, see PM33
WP4_KPI04_A	Course feedback	Rating of the training course based on the participant course feedback survey	60% responses "Good" or "Excellent"	80% responses "Good" or "Excellent"	91%	n/a, see PM21	94%	n/a, see PM27	100%	n/a, see PM33
WP4_KPI04_B	Course feedback - response percentage	Measure of how representative the collected feedback is	50% response rate for feedback survey	75% response rate for feedback survey	100%	n/a, see PM21	100%	n/a, see PM27	90%	n/a, see PM33
WP4_KPI05_A*	<i>Long-term impact, measured through post-6 month survey</i>	<i>Long-term perspective of the usefulness of the course</i>	<i>60% would recommend or has already recommended</i>	<i>90% would recommend or has already recommended</i>	<i>66.7% "Yes, I already have" and 33.3% said "Maybe".</i>	<i>n/a, see PM21</i>	<i>n/a, see PM21</i>	<i>n/a, see PM21</i>	<i>n/a, see PM21</i>	<i>92% either would recommend to others or already have</i>
WP4_KPI05_B	Course feedback - response percentage	Measure of how representative the collected feedback is	15% response rate for long-term feedback survey	20% response rate for long-term feedback survey	19.6	n/a, see PM21	n/a, see PM21	n/a, see PM21	n/a, see PM21	26% (6-12 months) & 17.5% (2 years)
WP4_KPI06_A	Presence in media	Raise awareness about our services to the wider community	* 1 publication per year in general / broad public journals	* 2 publications per year in general / broad public journals	0	0	0	0	0	0

D4.6 – Final report on dissemination and training

WP4_KPI06_B	Social media - Twitter	Regular engagement and updates are needed for social media channels	15 tweets per month	20 tweets per month	19.6	16.7	22	21	25	39
WP4_KPI06_C	Social media - LinkedIn	Regular engagement and updates are needed for social media channels	1 update per month	3 updates per month	1.3	1	1.7	0.6	0.3	0
WP4_KPI07	Number of website visits	This KPI helps to determine whether we are increasing our reach as well as when peak interaction periods are	600 per month	1200 per month	844	926	1002	1679	1392	1523
WP4_KPI09	Number of new Twitter followers	Similar to WP4_KPI08, determine whether our reach is increasing	15 new followers per month (i.e. up to 150% increase by PM36)	20 new followers per month (i.e. up to 200% increase by PM36)	19.7	17	36.7	35	39	67
WP4_KPI10	Number of unique subscribers to newsletters and mailing lists	Similar to WP4_KPI09, WP4_KPI10 determine whether our reach is increasing	715 by PM36 (i.e. 45% increase by PM36)	812 by PM36 (i.e. 65% increase by PM36)	530	557	673	897	934	1074
WP4_KPI11	Number of publications acknowledging the BioExcel project	Measurement of the impact the project has through publications	4 per year	8 per year	9	9	11	12	15	15

Notes:

WP4_KPI05_A*: Note that due to revisions of the long-term feedback survey the exact question (response: useful/not useful) was not asked in the survey. Originally envisage to be reported under WP4_KPI05_A is no longer in the survey template. Therefore, we report statistics on the survey question “Would you recommend this course to others