EXCELERATE Deliverable 10.4

Project Title:	ELIXIR-EXCELERATE: Fast-track ELIXIR implementation and drive early user exploitation across the life sciences	
Project Acronym:	ELIXIR-EXCELERATE	
Grant agreement no.:	676559	
	H2020-INFRADEV-2014-2015/H2020-INFRADEV-1-2015-1	
Deliverable title:	ELIXIR Advanced workshops on genome annotation	
WP No.	10	
Lead Beneficiary:	34 - UOCHB	
WP Title	ELIXIR Node Capacity Building and Communities of Practice	
Contractual delivery date:	30 June 2017	
Actual delivery date:	14 July 2017	
WP leader:	Jiří Vondrášek, Bengt Persson, Brane Leskosek	31: LiU
Partner(s) contributing to this deliverable:	SIB, MU, UOCHB, EBI, UTARTU, BSRCAF, CNIO, CSIC, FVIB, CSC, CNRS, HUJ, CNR, UiB, UiT, FCG, INESC-ID, LIU, UL,UOXF	

Authors and Contributors:

Henrik Lantz and Brane Leskošek

Table of contents

1. Executive Summary

Genome assembly and genome annotation constitute areas where significant resources, both manpower and computational resources, are needed to successfully conclude a project. Experience, both organism-specific but also general knowledge of assembly and annotation, is necessary and this is often slowly built up over the years when working continuously on similar projects. It is rare that the individual research group can find the time and resources necessary to build up this competence. Every project is unique, as all organisms are different and variety of different data types and tools exist. There are no golden standards, and the technology is rapidly advancing with new sequence technologies becoming available at a quick pace.

Experience in genome assembly and annotation exists in Europe but it varies between the ELIXIR Nodes, with some already having built up stable infrastructures and others just starting out. However, the need exists in all countries, not least because new sequencing technologies have made genome assembly and annotation more affordable and more feasible even over this past year. Consequently, one of the major goals of EXCELERATE task 10.3 is to build the capacity in genome assembly and annotation where it currently is insufficiently developed.

We report here on two workshops and two courses organized by the task in 2016 and 2017. We also briefly outline planned activities.

2. Impact

Activity	Location	Duration	ELIXIR nodes	ELIXIR staff	Students
Workshop	Barcelona, Spain	1 day	Belgium, EMBL-EBI, France, Portugal, Sweden, Switzerland	11	N/A
Workshop	The Hague, the Netherlands	½ day	Belgium, Czech Republic, EMBL-EBI, France, Slovenia, Spain, Sweden	9	N/A
Course	Prague, Czech Republic	3 days	Belgium, France, Norway, Portugal, Sweden	9	24
Course	Ljubljana, Slovenia	5 days	Belgium, France, Norway, Sweden	9	20

3. Project objectives

With this deliverable, the project has reached or the deliverable has contributed to the following objectives:

No.	Objective	Yes	No
1	Implement a programme of organisational capacity building in newly formed ELIXIR Nodes, including sharing of best practice between partners in accessing EU Structural Funds (ESIF) for operating infrastructure		x
2	Construct and coordinate ELIXIR-wide 'communities of practice' that support and develop the professionals who deliver advanced data and bioinformatics support and services in ELIXIR Nodes)	x	

4. Delivery and schedule

The delivery is delayed: ☑ Yes □ No

Deliverable has been delayed to include outcomes from the June workshop in Ljubljana 19-23 June 2017.

5. Adjustments made

N/A

6. Background information

Background information on this WP as originally indicated in the description of action (DoA) is included here for reference.

Work package number	10	Start date or starting event:	month 1
Work package title	**ELIXIR Node Capacity Building and Communities of Practice**		
Lead	Jiří Vondrášek (CZ) and Bengt Persson (SE)		

Participant number and person months per participant
1 – EMBL 6.00, 2 – UOXF 4.00, 5 – UTARTU 20.00, 7 – CNIO 1.00, 9 – CIPF 3.60, 13 – CSIC 2.00, 16 – FCG 2.00, 17 - INESC-ID 10.00, 20 – CSC 4.00, 21 – UiB 4.00, 23 – UiT 4.00, 26 – CNRS 5.00, 31 – LIU 24.00, 32 – UL 30.00, 34 – UOCHB 8.00, 35 – MU 26.60, 37 – VIB 10.00, 39 – BSRCAF 12.00, 40 – HUJ 8.00, 42 – FORTH 6.00

This WP will address the issue of how to get people in Nodes coming together in capacity building, as detailed in the tasks below. There will be accompanying training needs in this capacity building and those training needs will be addressed in WP11. The training needs are in advanced training of the staff handling data and performing genome annotation and assembly. Other training needs for Use Cases will be in general addressed in WP11, but not specific to every Node. For Node capacity building, advanced training will be needed also in management and know-how on operating Nodes, performed in close collaboration with Task 10.1.

A Community of practice is a group of people who share a craft or a profession, created to coordinate efforts to solve defined tasks and/or with the goal of gaining knowledge related to their field. ELIXIR is looking to establish such Communities of Practice of bioinformatics experts involved in advanced bioinformatics user support across the Nodes to effectively interact with bioinformatics infrastructure users at interfaces of different research fields. ELIXIR Communities of Practice would be the primary mechanism for ELIXIR to establish domain specific services, for example, forming a community of genome annotators across Nodes to meet the need from national researchers of ready access to genome annotation resources. Other examples could be to meet the needs of Rare Disease or Medical genomics research, agricultural or marine bioinformatics and chemical compounds for biology.

ELIXIR will start to build these Communities of Practice to enable coordination and knowledge exchange in selected areas in tasks 10.2 and 10.3. Task 10.2 is directed to create Good Practices in setting up data Nodes, of importance to create a sustainable and scalable data flow from laboratories to national Nodes and further to European or global databases. Task 10.3 is directed to coordinate and exchange expertise in the field of genome annotation and assembly and to create Good Practice in for this field. In the future, further communities of practice are envisioned, arising from needs identified by the Use Cases (WP6 to 9) and identified through community workshops and surveys (Task 10.4). The creation of a sustainable mechanism for establishment of communities of practice is also addressed in Task 10.4.

Objectives

WP10 is focused on strengthening the ELIXIR infrastructure by supporting coordination of Node activities and increasing the organisational capacities of ELIXIR Nodes. ELIXIR Nodes are at very different levels of maturity, ranging from national infrastructures that have existed for over a decade to newly formed consortia. Activities will focus on spreading the knowledge and bioinformatics best practice that exists within ELIXIR's larger and more established Nodes, with newer or smaller ELIXIR Nodes in less research-intensive areas of the EU. This will help to create a stairway to excellence for partners involved, and support the creation of a true European Research Area. One of the deliverables will be a set of "Good practices" for setting up and running an ELIXIR Node, which will be of substantial value for both current and future Nodes.

Its two Objectives are:

1. Implement a programme of organisational capacity building in newly formed ELIXIR Nodes, including sharing of best practice between partners in accessing EU Structural Funds (ESIF) for operating infrastructure.
2. Construct and coordinate ELIXIR-wide 'communities of practice' that support and develop the professionals who deliver advanced data and bioinformatics support and services in ELIXIR Nodes.

Work Package Leads: Jiří Vondrášek (CZ) and Bengt Persson (SE)

Description of work and role of partners

Task 10.1: ELIXIR Node Capacity Building (46PM)

This task will support the formation of an ELIXIR community. There are significant differences between existing ELIXIR Nodes in their capacity, level of expertise and maturity of services/tools/data. We will increase the joint competence and capacity for Nodes lacking a large national user community, large-scale projects and big data or having a limited record of offered tools and services. These Nodes will benefit from mutual collaboration and connection with well-established and more advanced Nodes they can utilize their know-how for a more rapid Node development. Altogether, this will help shape ELIXIR as an efficient pan-European infrastructure.

The major aim of this task is to provide management knowledge transfer among Nodes to create a set of well-balanced, well-functioning and compatible Nodes. Support in coordinating national Nodes, including Skills and Knowledge exchange between ELIXIR Nodes. Nodes with different experiences will help to provide knowledge regarding good practice in different situations and providing direct support to implementation of national infrastructures (e.g. by national / regional workshops with external experts, support to national community building efforts). The heterogeneity of Nodes established will help providing multiple effective ways for coordination and to get funds from national providers and their commitments to the infrastructure. Knowledge exchange will be catalysed by workshops, staff exchange programme and visits. This activity is based on the ELIXIR community practice experience but it is more general and should cover some features brought

by larger staff community.
Identify and apply technical solutions at/between Nodes. The reason for particular technical solution must be explicitly formulated and the solution must be applicable on more than 2 Nodes. The capacity building deliverables would be primarily workshops based on Technical Services and/or Training WP deliverables.
Partners: CH, CZ, EE, NO, PT, SE, SI, UK, ES, EL, IL, EMBL-EBI

Task 10.2: Capacity Building in Data Nodes Network (34PM)
One of the aims of ELIXIR is to establish a network of data Nodes (Nodes with large data collections and databases with established way of data deposition and curation) to enable scalable data storage and their transferability by means of standardised formats. In this task, we will focus on establishing guidelines and good practices to facilitate efficient data collection into core data resources (cf. WP3), primarily focusing on data needed for selected Use Cases (WP6 to 9). This is tightly linked with IT solution by means of storage, dedicated networks and connections (cf. WP4). A distributed network following the same standards will also simplify international sharing of datasets for which this is ethically permitted.
This task both includes creation of routes for data publishing in a uniform manner across ELIXIR with data Nodes in each country and includes data repositories for replication of reference data allowing for fast access. The setting up of a data Nodes network has been identified by the technical experts within ELIXIR as a prioritised area.
Task 10.2 also includes development of Good Practices in setting up data Nodes enabling secure storage of sensitive data, such as sequence data related to patients. The task is interfacing with WP4 regarding technical developments on AAI and data transfer. Furthermore, there are connections with WP4 on data interoperability and the Use Case in WP9 on sensitive data.
Partners: SE, FI, CZ, EMBL-EBI, SI, PT, ES, EE. In due time, all ELIXIR Nodes are expected to have an ELIXIR data Node.

Task 10.3 – Capacity Building in Genome Assembly and Annotation (44PM)
Specialised expert platforms for genome assembly and annotation are already available in several ELIXIR countries. They provide critical support to complex genome projects and deliver annotations that serve as the basis for scientific inquiry into the genomics of newly sequenced organisms. The specialised expertise at multiple ELIXIR Nodes would benefit from capacity building through competence-spreading advanced workshops and staff exchange.
The capacity-building efforts will benefit the Use Cases in WP6 on marine organisms and in WP8 on plant Use Cases. The genome annotation groups will contribute with domain-specific knowledge about different species, e.g. marine organisms (SE, NO), woody plants (PT) and crop plants (SI).
Furthermore, in order to facilitate access to genome annotation to the users, we propose a deployment of web services to enable genome projects in the scientific community to efficiently interact with the data. The development of such web services is intended together with the EnsEMBL team to create a pan-European collaboration on genomics resources to provide researchers with a unified analysis platform carried by multiple partners.
Partners: SE, NO, FR, PT, EBI, SI, BE, CZ, ES.

Task 10.4 – Sustainability of capacity building (30PM)
The main goal of Task 10.4 is periodical and long-term discovery of users with specific capacity needs at ELIXIR Nodes and/or research groups within Nodes. This knowledge of capacity needs/gaps will be gathered through surveys and face- to-face meetings. With capacity needs identified the Task 10.4 team will connect users with WP11 groups that have at their disposal training infrastructure, learning materials and knowledge needed to implement the capacity building. In order to ensure the sustainable flow of knowledge and stable capacity maintenance we need to provide long-term networking of capacity seekers and providers. They will be focused to the great extent to the Good Practices from Task 10.2 and 10.3 (and WP6 to 9). With well-formed ELIXIR Communities of practice, the Task 6.4 will be able to lead the reuse or even suggest the adaptation of WP11 courses and training materials for specific capacity building needs.
It is of great importance that capacity needs will be periodically (but in long-term perspective) tested through surveys, which will also contribute to the sustainability of training infrastructure and learning materials provided by WP11. Task 10.4 will monitor the implementation of capacity building in Tasks 10.1, 10.2 and 10.3 in order to extract good practices and compile good practice recommendations and guidelines which can be used in other capacity building contexts.
Partners: SE, SI, CZ, BE, EE, EL, IL, EMBL-EBI

Task 10.5: Supporting ELIXIR Nodes in understanding Smart Specialisation Strategies and accessing EU Structural and Investment Funds (ESIF) (36.2PM)
The potential for exploiting funding synergies between EU Research programmes and ESIF are well known62. Those ELIXIR Nodes eligible for ESIF are therefore presented with a real opportunity for local funding of their Node, particularly in light of the proposed focus on ESIF and ESFRI that many Member States are making within their national plans to the Junker Investment Plan. However, understanding the local priorities for funding, rules, and application procedures presents is complex and time consuming and securing ESIF for operational costs of life science infrastructures is a real challenge. For ELIXIR Nodes to access ESIF in any meaningful way, support needs to be targeted at the local level, allowing scientists to build up an understanding of their local Smart Specialisation Strategy, which dictates the funding opportunities for that region, and then develop a strong business case that can be used for subsequent funding applications.
Partners: CZ, SI, EE, EL

ELIXIR ESIF Task Force (Months 1-12)
ELIXIR Structural Funds Task Force grouping funding specialists across ELIXIR Nodes will be established to share best practice in ESIF use for research infrastructures. The Task Force would also engage external experts such as ones from national managing authorities for ESIF, DG REGIO, DG EMPLOY, DG Enterprise and Industry and Jaspers and would make use of existing reports such as the ESPON KIT report (www.espon.eu).
An ELIXIR-wide Workshop early at start of the project to pool good practice on using Structural Funds to support research infrastructures and facilitate personal interactions. Meeting will be hosted and organised by CEITEC, who leads this task.

This would include talks from ELIXIR Nodes with experience of accessing Structural Funds (Estonia, Czech Rep, Slovenia), as well as other ESFRIs such as ELI that have done this successfully in other disciplines
Local priorities and their overlaps identification towards Business Case (Months 6-24)
As all regional priorities are different, and as the application process for funding is done in the local language and following local rules, target Nodes will work with their regional partners to understand the priorities. This task will support Nodes in understanding their local Smart Specialisation Strategy and the regional priorities relating to research and life sciences. Access support from Jaspers following the connections built up within Months 1-12.
Supporting Nodes in actually developing the Business Cases and applications for Structural funds to support the construction and/or operation of the Node. The timing of this work will depend on when the calls will be opened for each region.
Partners: CZ, SI, EE, EL

7. Appendix 1: Report on workshops and courses in genome assembly and annotation

ELIXIR-EXCELERATE Deliverable 10.4 – Advanced Workshops in Genome Annotation

Summary

In this document we report on four workshops arranged in 2016 and 2017 by ELIXIR-EXCELERATE task 10.3 ”Capacity building in Genome Assembly and Annotation”. The first two were used to plan and structure the activities of the task, while the others were full courses held in the Czech Republic and in Slovenia to increase the capacity in genome assembly and annotation in those two countries. The response to the courses has been very positive, with an average of 4.68 and 4.65 resp. in general rating (max 5.0). 100% of the participants who answered the survey would recommend the course to someone else. We also outline planned activities for the task and discuss improvements of our current efforts. We conclude that there is a demand for our activities and that we are filling an important niche.

Report

Table 1. Activities covered in this report

Activity	Location	Date
Workshop	Barcelona, Spain	April 3, 2016
Workshop	The Hague, the Netherlands	Sep 4, 2016
Course	Prague, Czech Republic	Oct 12-14, 2016
Course	Ljubljana, Slovenia	June 19-23, 2017

Nodes financed in the task are Belgium, Czech Republic, France, Norway, Portugal, Slovenia, Spain, and Sweden.

Already reported activities

The first workshop of the task was held in in Uppsala, Sweden, in October 2015, and this has already been described in Deliverable 10.3. The main purpose of this workshop was to establish contact between the nodes to enable a Community of Practice, and make a list of available expertise for the involved nodes. Feasibility of different areas, such as development and courses, was also discussed.

Workshop in Barcelona, April 3, 2016

Table 2. Barcelona workshop, general information

Duration of workshop	Nodes represented	Number of attendees
1 day	Belgium, EMBL-EBI, France, Portugal, Sweden, Switzerland	11

A workshop in genome assembly and annotation was held in conjunction with the ISCB conference in Barcelona on April 3, 2016. The conference focused this year on genome annotation and it was therefore judged beneficial to have a workshop at the conference to be able to attract more researchers interested in the topic. Eleven researchers representing 6 Elixir nodes were present at the meeting.

All nodes were asked to present their activities and interest in the genome assembly and/or genome annotation. The discussion then focused on three areas: Network of experts, training, and good practices in genome assembly and annotation.

Network of experts – Assembly and annotation experts are present in many European countries, but are often isolated from each other. To enable communication between these experts, they must be made aware of each other and there must exist a means of communication between them. This Network of Experts could then be utilised to help nodes interested in increasing their capacity in the subject and to spread advanced knowledge in Europe.

Several technical solutions were discussed and the chat and messaging service Slack was eventually chosen as the best option. A slack instance for the task was also created after the workshop.

Training – It was generally agreed that training was needed to increase the capacity in genome assembly and annotation in Europe. Several nodes expressed interest in teaching at courses, but "immersive training" where researchers would stay for shorter periods at expert centres in Europe was also mentioned as an interesting possibility. It was decided to continue the discussion at the next workshop planned for October in The Hague.

Good practices – Nodes interested in increasing their capacity expressed an interest in documents of Good Practices. These documents could be used to ease the single

researcher into the subject but also be used as an aid for nodes interested in setting up their own core facilities. It was decided to focus the next workshop in The Hague on "Good Practices in Genome Assembly and Annotation".

Workshop in The Hague, Sep 4, 2016

Table 3. The Hague workshop, general information.

Duration of workshop	Nodes represented	Number of attendees
1/2 day	Belgium, Czech Republic, EMBL-EBI, France, Slovenia, Spain, Sweden	9

This workshop was held in conjunction with the ECCB meeting in The Hague on Sep 4, 2016. The focus was on Good Practices, but time was also set aside to discuss planned training activities.

Good Practices – Already at the Barcelona workshop it was decided to create documents for good practices in genome assembly and annotation. What was now needed was to decide what the document should include and how it should be published.

Based on the fact that sequencing technologies are rapidly changing, it was clear that a written publication easily could become outdated if it focused too much on specific tools. It was therefore decided to have more general information in the publication, but couple this document to an online WIKI which could include tool-specific information, and be more easily maintained and updated. Faculty1000 was suggested as a potential journal for the written publication. The target audience was suggested to be individual researcher rather than core facilities, although parts of the information of course could be of value to those facilities too.

The scope of the document and WIKI was also discussed. All sequencing technologies of relevance to genome assembly or annotation should be covered, and how to deal with these in assembly or annotation. Other sections of importance discussed were DNA and RNA quality, sequence submission, structural and functional annotation, and formats. It was decided not to include metagenomics. The WIKI would include information of specific tools as well as suggestions of workflows. It was decided to maintain the WIKI during the EXCELERATE funding period, and then take a decision to continue or not.

Course in Prague, Czech Republic – During summer it was suggested to have a course in the Czech Republic during autumn and this was now decided for October. A Bring Your Own Data approach was suggested, but was decided to be impractical for a three-day course as many analyses easily can run for days. The French node would prepare virtual machines in collaboration with the Czech and Slovenian nodes. The importance of gathering information on the participants´ interests in course

registration was stressed to allow us to tailor the course after the participants' needs. The eLearning platform (EeLP) developed by ELIXIR-SI was decided to be used for hosting documents and also a survey after the course.

Course Prague Oct 12-14, 2016

Table 4. Prague course, general information.

Duration	Nodes represented	Number of teachers	Number of students
3 days	Belgium, France, Norway, Portugal, Sweden	9	24

All students were from the Czech Republic, PhD-students or more senior.

Description incl. technical details – Two days were focused on genome assembly and one day on genome annotation. When registering for the course the students were asked to supply information on their research and if they were interested in any particular subject. This allowed us to tailor the course based on their interests. The course included a mix of lectures and practical exercises using Linux command line. Prior knowledge of Linux command line was a requirement for the course and the students were asked to state their level of expertise when registering. We also included a short session for discussion of the participants' own projects. In this session the participants were split into smaller groups with one or two teachers per group to allow teachers to give direct feedback on the participants' research projects. The ELIXIR eLearning platform developed by the Slovenian node was used to host presentations and exercises and a survey used to evaluate the course. A working dinner was held on the second day to get a chance to socialize more with students and collect more information about the type of support they needed in their research.

Two Virtual Machines were created for the course; one by the French node and one by the Czech node. Both of the VMs were hosted by the Czech node.

Outcome –

Table 5. Selected answers from the survey (19 out of 24 participants answered the survey, presented is the average of answers, max 5.0)

How would you rate the workshop in general?	Did the workshop meet your expectations?	Would you recommend the course?
4.68	4.42	100% yes

The course was in general very much appreciated by the students. The long reads lecture was the single most popular part (average 4.89) and only one lecture received

an average below 4 (3.89, transposable elements), all other lectures and exercises were rated 4.21 or higher. The students clearly appreciated the chance to talk to experienced bioinformaticians and during the dinner several participants expressed that they were mostly limited by access to knowledge rather than compute resources or access to data. Several participants also expressed a need for more time during the course, they felt that did not have time to go through all exercises and needed more time to digest the information.

We feel that some factors in particular can explain the in general very high ratings we received. These are: a high ratio of teachers to students, teachers with a very high level of expertise, a high level of command line Linux among the students, and the fact that the course had a clear subject and thus attracted students already interested in the subject.

All participants were invited to the already existing Slack channel set up for the task to facilitate communication with the teachers after the course. All material in the eLearning platform will also continue to be available.

Changes to future courses based on survey – Two follow-up meetings were held via Skype with the involved teachers to discuss how the course could be improved. Two main improvements were suggested: 1) The course should be extended to 5 days to allow for more time for in particular the exercises. 2) The students clearly appreciated the chance to get feedback on their own research. We should therefore increase the time spent on discussion of the participants' own projects and also facilitate for them to bring their own data to analyse together with the teachers.

Work was also started by the Czech node to merge the two existing VMs into one.

Course Slovenia, June 19-23, 2017

Table 6. Slovenia course, general information.

Duration	Nodes represented	Number of teachers	Number of students
5 days	Belgium, France, Norway, Sweden	9	20

One student was from Belgium, all the others were from Slovenia. All were of PhD-student level or more senior. The number of students was limited to 20 to provide optimal interaction with teachers and enough compute power for VMs. The interest for participating the course was thus much higher.

Description incl. technical details – Based on feedback from the previous course, this course was extended to 5 days. Two days were spent on genome assembly, two days on genome annotation, and a full day of discussion and work with the participants'

own data in between those two parts. Very little content was added compared to the Prague course, the idea was to simply give them more time to finish, and also understand, the material. Only two things were added: an exercise on transposable elements and a bacterial annotation part. Bacterial annotation was added as a direct response to the information received in the registration as many participants stated they were working with bacteria. All teaching material and the course survey was also this time hosted in the ELIXIR eLearning platform (EeLP).

Due to technical challenges, we ended up having three Virtual Machines for the course. Two of them were hosted by the Slovenian node, one by the French node.

Outcome –

Table 7. Selected answers from the survey (17 out of 20 participants answered the survey, presented is the average of answers, max 5.0)

How would you rate the workshop in general?	Did the workshop meet your expectations?	Would you recommend the course?
4.65	4.47	100% yes

The questions in the survey were updated and are now standardised and aligned with the ELIXIR TrC standard set of questions for short term feedback and the results will be included in the metrics report of the ELIXIR Training platform. The course was very successful, getting overall good ratings, comparable to or better than the results from the Prague course. The highest individual rating (average rating 4.88) was given to the introductory lecture, and all parts received ratings of 4.0 or higher. The discussion day was popular (average rating 4.71) with three participants also mentioning it as the best part of the course. The technical issues with the Virtual Machines (VMs) was the largest cause for frustration. To avoid this next time and to join the necessary tools into one VM we started to prepare the VMs guidelines that will be used for next courses.

In comparison to the Prague course, the average Linux expertise of the students was lower even though the knowledge of Linux command line was a requirement for the course. This made progress slower, with some students struggling to get through the exercises. As many teachers and local technical help were present in the classroom at all times, quick help was always possible, but for future courses the knowledge of Linux command line should be assessed in the prologue of the course. The sample assessment quiz will be prepared in the eLearning platform (EeLP) for the next course.

All participants were invited to the Slack to facilitate communication with the teachers after the course. All material in the eLearning platform will also continue to be available.

Changes to future courses based on survey and experiences from the course –

A short primer on Linux command line, just to refresh the memory, and self-assessment quiz for Linux command line knowledge will be included at the prologue of the course. The Virtual machines also need to be merged, and work has already started to document hardware requirements and tool specific dependencies that will ease setup for the next course. Access to the VMs will be better embedded into the eLearning platform (EeLP).

Planned activities

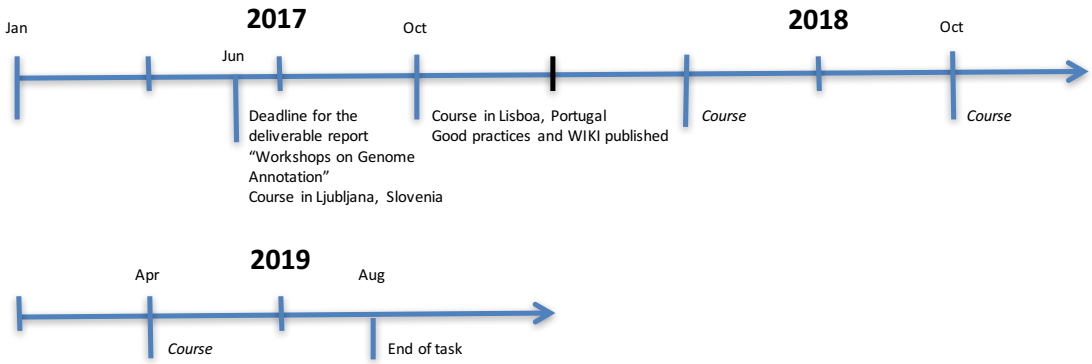

Fig 1. Timeline of present and planned activities. 2018 and 2019 activities tentative.

Courses – A third course, this time also five days, is planned in Lisboa, Portugal, 23-27 October. After this we would like to continue to host courses 2 times per year, but no specifics are yet decided.

Good practices – A printed publication and an accompanying WIKI is preliminarily planned to be released in September 2017. The printed publication will be more general and more stable over time, while the WIKI will be more detailed and go into specific tools. A draft manuscript exists, and the WIKI is also in place (managed by the French node) and we are now in the middle of filling the WIKI with information. Interlinking between WIKI and materials in EeLP will be established and maintained.

Project support – We are currently investigating the possibility of having the more experienced nodes offer support with analyses in genome assembly and annotation. A pilot project with the Swedish, Czech and/or the Slovenian nodes is being discussed, and should start in in 2017.

Immersive training – Several nodes are prepared to host researchers from nodes interested in learning more of genome assembly and annotation. There is an interest among researchers in the Czech Republic and Slovenia to take part of such training, but they are currently limited by available funding.

Conclusion

Based on feedback from in particular our courses, we can conclude that ELIXIR-EXCELERATE task 10.3 “Capacity Building in Genome Assembly and Annotation” fills a need in Europe. Our overall rating for the courses are 4.68 and 4.65, which out of a maximum of 5.0 can be considered very high. Researchers in countries where the available bioinformatics knowledge is scarce feel limited by their access to knowledge, and greatly appreciate the possibility to get in touch with experienced bioinformaticians through our courses and other events.

A core set of experts and means of communication between these experts is now established. What is now needed are more efforts to make the capacity building sustainable, so that the knowledge continues to live in the countries we are in touch with.