Recommendations for a coordinated approach to bioinformatics user training in Europe

ELIXIR’s training strategy committee

- Anna Tramontano (chair), CRS4, Pula, Sardinia, Italy
- Cath Brooksbank (co-chair), EMBL-EBI, Hinxton, Cambridge, UK
- Vicky Schneider (project manager), EMBL-EBI, Hinxton, Cambridge, UK
- Thomas Blicher, Technical University of Denmark, Lyngby, Denmark
- Tommi Nyronen, CSC - Scientific Computing Ltd., Espoo, Finland
- Heta Kero, CSC - Scientific Computing Ltd., Espoo, Finland
- Kari Tuononen, University of Helsinki, Finland
- Hans Werner Mewes, German National Center for Health and Environment, Neuherberg, Germany
- Janusz Bujnicki, International Institute of Molecular and Cell Biology, Warsaw, Poland
- Jaak Vilo, University of Tartu, Tartu, Estonia
- Pedro Fernandes, Instituto Gulbenkian de Ciencia, Oeiras, Portugal
- Terri Attwood, University of Manchester, UK
- Matthias Haury, EMBL, Heidelberg, Germany

We warmly welcome feedback from ELIXIR’s stakeholders. Please address your comments to elixir-wp11@ebi.ac.uk
Contents

ELIXIR’s training strategy committee ........................................................................................................... 1
1. Executive summary ...................................................................................................................................... 3
   1.1 Summary of this report ....................................................................................................................... 3
   1.2 Summary of recommendations .............................................................................................................. 4
2. Scope of this report ...................................................................................................................................... 5
3. Outline rationale ........................................................................................................................................ 6
   3.1 Potential opportunities and benefits ...................................................................................................... 6
   3.2 Stakeholders overview ......................................................................................................................... 8
   3.3 Needs and barriers to progress ............................................................................................................ 10
   3.4 Challenges and risks ........................................................................................................................... 11
4. Description of roadmap in terms of elements/key themes ........................................................................ 13
   4.1. State of the art .................................................................................................................................. 13
   4.2. Specific opportunities ......................................................................................................................... 21
   4.3. Specific barriers to progress .............................................................................................................. 22
   4.4. Specific challenges to be addressed .................................................................................................... 22
   4.4.3 The ELIXIR training support unit .................................................................................................... 24
   4.5. Priorities mapped to time ................................................................................................................... 29
   4.6. Specific dependencies ....................................................................................................................... 31
5. Benefits of the ELIXIR training infrastructure ...................................................................................... 31
   5.1. Contribution to European capacity building ...................................................................................... 31
   5.2. Contributions to and synergies with national scientific activities/internal roadmaps in member states .................................................................................................................................................................................. 34
   5.3. Contributions to and synergies with other ESFRI projects ............................................................... 35
   5.4. Potential benefits to stakeholder communities ................................................................................. 36
6. Steps for realising the benefits .................................................................................................................. 36
Appendix 1: Examples of Current bioinformatics training programmes ......................................................... 38
   e-learning ............................................................................................................................................... 41
Appendix 2: thoughts on benchmarking ........................................................................................................ 44
Appendix 3: thoughts on monitoring success of training courses .................................................................... 45
1. Executive summary

1.1 Summary of this report

- Elixir is concerned with the design of a sustainable infrastructure for biological information in Europe to support life science research and its translation to medicine and the environment, the bio-industries and society. Whilst ELIXIR’s broad aim is to improve access to biological data for Europe’s life science community, the goal of ELIXIR’s training strategy is to improve accessibility, by empowering European researchers to make effective use of the data.

- ELIXIR’s training strategy is tightly focused on supporting the provision of user training for the databases and tools that will be provided by ELIXIR; the main beneficiaries will be users of these resources. We do not specifically address bioinformatics education (undergraduate, postgraduate and postdoctoral training in bioinformatics), as we consider this to be the role of universities. Nevertheless, ELIXIR’s training strategy will make teaching materials more readily available to all, and will therefore be of direct relevance to bioinformatics education.

- This report is of interest to:
  o Biological data providers
  o End-users of bioinformatics resources
  o Power users of bioinformatics resources
  o Trainers and educators in the life sciences
  o Biological data resource and tool providers
  o Funders of infrastructure

- Whilst there is already a significant amount of bioinformatics user training ongoing in Europe, there is little or no coordination among training centres of excellence. This has led to imbalances, both in terms of the geographical distribution of training and in terms of the subjects taught. ELIXIR’s training strategy aims to address this fragmentation, to support the provision of a high standard of user training to Europe’s entire life-science research population.

- Current barriers to progress include:
  o The rapid evolution of bioinformatics resources, which makes the provision of up-to-date training materials extremely challenging;
  o The lack of a centralised body to which stakeholders in bioinformatics user training can turn for guidance;
  o Lack of recognition of the importance of bioinformatics user training, even within the bioinformatics community.

- ELIXIR proposes to overcome these barriers by:
  o Creating a mechanism by which the development of data resources is tightly linked to the provision of training materials;
  o Creating a training support unit to: (a) provide a centralised training registry that will improve access to bioinformatics user training throughout Europe; (b) support trainers throughout Europe by providing access to regularly updated training materials that can be customised to create courses that are directly relevant to their local trainees; (c) develop benchmarking and evaluation systems; (d) provide mechanisms for developing, piloting and evaluating new training programmes; and (e) act
as a single point of contact for national and pan-European training infrastructures and projects.

- Throughout this report, we make a clear distinction between training materials and courses. ELIXIR’s training infrastructure will focus on improving access to training materials that can be adapted by trainers as appropriate for their own purposes. Rather than developing a ‘one size fits all’ approach to user training throughout Europe, we aim to create resources that support trainers and that help trainees to find out what training is available and which courses best suit their needs.

- ELIXIR will adopt a phased approach to building its training infrastructure, beginning with a comprehensive survey of current user training activities, making this information openly available for all and building a community of stakeholders. Consultation with this community will help to drive improvements in the provision of user training throughout Europe.

- Failure to invest in a proper user-training infrastructure will ensure the failure of ELIXIR by preventing Europe’s researchers from unlocking the enormous potential of our public biological data. This will have a negative impact not only on academic research, but also on industry; furthermore it will prevent the newer EU member states from realising the benefits of ELIXIR, by perpetuating uneven access to bioinformatics user training.

- The cost of a training infrastructure for Europe’s core biological data resources and tools is small compared with the cost of collecting, storing and organising the data. This relatively small investment will make a significant difference to the effectiveness of ELIXIR by ensuring that its users are empowered to make the most of the data. This will enable funders to realise the full benefit of their investment in bioinformatics infrastructure.

- If adequately resourced and properly managed, ELIXIR’s training strategy will make the benefits of ELIXIR accessible and broadly applicable to Europe’s entire research community. It will permit the exploitation of Europe’s uniquely rich data collection throughout the academic and commercial sectors, permitting the translation of biological data into discoveries that will improve quality of life for Europe’s citizens and increase Europe’s competitiveness on a global scale.

1.2 Summary of recommendations

ELIXIR’s training committee makes the following recommendations for the creation of a sustainable bioinformatics user-training infrastructure in Europe. These recommendations are detailed further in section 4.4 and are summarized in Fig. 8.

Training must be tightly integrated with the development of ELIXIR’s data resources and tools to keep pace with changes to these. To enable this, we recommend:

- Funding for adequate training/outreach staff as an integral part of the funding mechanism for ELIXIR data resources and tools.

- With each release of an ELIXIR resource, a standardised data sheet is provided to facilitate the creation and maintenance of training materials.

- Each resource develops and updates a small set of standardised use cases so that they remain robust with each release cycle.

- Data resource providers use an early warning system to alert trainers to any major updates to a resource that will have knock-on effects on user training.
To make bioinformatics user training accessible to ELIXIR’s rapidly growing and diversifying user community, some level of central coordination is required. We recommend that a centralised Training Support Unit be funded to:

- Build and maintain a registry of information relating to bioinformatics user training. This will allow trainers to share and rate training materials and expertise with other trainers; it will allow trainees to find courses that are best suited to their needs.
- Develop benchmarking and evaluation systems.
- Coordinate support and developmental opportunities for trainers.
- Enable the development of new training programmes and their integration with existing initiatives.
- Act as a central point for interaction with other pan-European training infrastructures.

2. Scope of this report

Until very recently, the acquisition of data through experimentation was the major bottleneck to research in the life sciences. The recent explosion in high-throughput methodologies, most recently illustrated by the development of ever-faster and cheaper DNA sequencing technologies, has radically changed this situation, so that data analysis, as opposed to data acquisition, is now the greatest challenge. Enabling Life scientists to gain a firm grasp of the available bioinformatics tools and resources, and to understand the capabilities and limitations of these resources, will therefore be of utmost importance to Europe’s strength in life sciences research for the foreseeable future. ELIXIR must therefore consider how it will facilitate this process. We address here the issue of best practice principles for user training. It is not within the scope of this report to discuss education at the academic level, although it is apparent that basic courses in bioinformatics and computational biology should be included in all life science study courses and in computationally oriented curricula as well. We recognise that modern life scientists need to be educated in the new emerging areas of biology and medicine, and that this is a task for academic institutions. In spite of the differences in the education systems throughout Europe, the quality of the education in the area is excellent and sufficiently homogeneous to allow scientists to communicate and collaborate, effectively taking advantage of a similar scientific background. Less ideal is the situation as far as bioinformatics user training is concerned. The tools are many, often offered in different flavours and with technical differences that are not easy to appreciate for an experimental biologist, for example. Development of services is, commendably, a very widespread activity and several laboratories offer resources, often but not always together with appropriate training material. Furthermore, although several laboratories and institutions offer courses at different levels, of varying length and user commitment, on a large variety of topics, there is little or no coordination or shared language among providers of such user training. This is very confusing for users, who are usually uncertain about the level of basic knowledge needed to make the best of the training resources. We will therefore limit this report to user training for data resources and tools that will form part of the ELIXIR infrastructure, and will outline our vision for how the user training landscape in Europe can be improved for the benefit of its life science community.

Although the core aim of ELIXIR is to serve European life scientists, it should also be borne in mind that bioinformatics resources are worldwide resources; a sustainable training infrastructure in Europe therefore has to take into account the need to
integrate with similar efforts on other continents. We plan to circulate this document to relevant institutions outside Europe and to create effective links and synergies with them.

The aim of this work package is to:

• Survey the state of the art in Europe as far as bioinformatics training is concerned (section 4.1);
• Perform a gap analysis of training modules essential for the correct use of the ELIXIR infrastructure, taking into account (1) gaps in training for particular subject areas and (2) geographical gaps in training provision;
• Make recommendations to developers of resources on basic information that they should provide to users and to trainers to facilitate the development of user training and to help users of ELIXIR’s data resources select appropriate courses (section 4.4.1.1);
• Make recommendations for a coordinated approach to bioinformatics user training in Europe (section 4.4.1.2);

Throughout this report, we make clear distinctions between training material, training ‘modules’ and training courses. We define these as follows:

**Training material** is any content that can be used for training purposes. It includes pre-structured content such as tutorials, worked examples, quizzes, and presentation materials such as slide stacks. Material that requires further structuring (for example, a review article about a data resource that could be used as the basis of a tutorial) can also be considered to be ‘raw’ training material.

**A training module** is a self-contained, re-usable block of training that can be used as part of a course. A module may have prerequisites associated with it (for example, it may be advisable to have completed an introductory module on a particular data resource before attempting more advanced training).

**A training course** is typically built of a selection of training modules. Our definition of training course includes both face-to-face training and distance learning.

3. Outline rationale

3.1 Potential opportunities and benefits

Bioinformatics is a core capability whose technology base firmly underpins 21st century biology. As a discipline, it is still evolving, bringing with it new technical problems and requiring new skills at all levels. Today, it is poised to provide new insights into critical medical, agricultural and scientific problems; but this potential will only be realised if scientists and clinicians have the necessary skills and support to make the best use of the ever-growing mountain of ‘omics’ data now available. Acquiring such skills is made harder by the rapid evolution of the field, the diversity of techniques within the field, the general lack of standards, and the difficulty in discovering suitable training opportunities - it is not so much that courses don’t exist, but rather, the problems that new-comers often experience in finding out about those that do. ELIXIR provides an opportunity to radically simplify access to information on bioinformatics courses and course materials. This would immediately benefit users and trainers.

Perhaps a more invidious problem arises from the wide variation in the degree of affluence across European countries. Less well-developed regions of the Community are unlikely to have critical mass in terms of expertise and personnel across the
spectrum of specialties that now fall under the banner of bioinformatics, making it harder to embrace this evolving discipline effectively. Often, a single individual takes on the mantle of ‘local bioinformatics expert’, having to trouble-shoot for bench biologists, to learn new skills as the need arises, and subsequently to pass on those skills to colleagues and students (often in an ad hoc fashion). Even when relevant training opportunities do exist in neighbouring countries, such organisations have insufficient funds to send groups of researchers or students to attend them. By bringing together stakeholders from throughout Europe, ELIXIR provides the opportunity to perform a thorough gap analysis of bioinformatics training in Europe, and to develop a strategy for filling these gaps. The very nature of bioinformatics makes it an excellent test case for developing and testing e-learning methodologies which, if successful, could remove many of the socioeconomic barriers to user training.

Providing professional, coordinated and effective training for users of ELIXIR’s databases and tools will improve Europe’s competitiveness on many levels, extending far beyond the benefits to those receiving the training. Such benefits include:

• More effective use of Europe’s biological data resources by life scientists, with a positive impact on European research output.

• A more uniformly high understanding (through the sharing of common training materials) and uptake of Europe’s bioinformatics resources, both within and beyond Europe.

• Uptake of the resources by user groups who have not traditionally used bioinformatics; for example, clinical geneticists and environmental researchers. In this respect, a bioinformatics training infrastructure for Europe has the potential to benefit not just ELIXIR’s stakeholders but also those of the other ESFRI BMS projects, the Innovative Medicines Initiative, some ESFRI environmental science infrastructures (e.g. LifeWatch) and other nascent ESFRI infrastructures with a requirement for biological data. This will provide an important route by which biological data is applied to medicine, agriculture and environmental science.

• Feedback from users, via trainers, will go back to the developers, leading to more rapid evolution of data resources and tools in line with users’ needs. This will help to break down the disciplinary boundaries between computational and experimental research, thereby helping to accelerate research progress.
3.2 Stakeholders overview

As outlined above, user training is of crucial importance to the delivery of ELIXIR’s goals. It is therefore relevant to all of ELIXIR’s stakeholder groups. We categorise these stakeholders as follows.

3.2.1 Data providers and users

The immediate beneficiaries of bioinformatics user training are those who need to make effective use of the data resources, tools and standards that come under the umbrella of ELIXIR. This user base is growing, not just in terms of numbers but also in terms of diversity as bioinformatics forges links with other disciplines such as medicine, chemistry, and environmental science. There is a spectrum of users with a correspondingly diverse range of training needs. For the sake of simplicity, we divide these into three broad categories, and recommends that the ELIXIR training infrastructure considers the needs of all three types:

(1) End users

We define end users as the people whom our data resources are designed to serve. Typically, these are experimental researchers in the basic life sciences, but this user group is now expanding to encompass a wide range of different types of professional, some not directly involved in research: for example, scientists such as cytogeneticists and clinical geneticists increasingly need to make use of human genome resources such as Ensembl to facilitate their diagnostic work. Bioinformatics-related tasks that end users may need to perform include (but are not limited to) the following:

- Searching and reviewing the scientific literature.
- Searching for individual genes or proteins, usually with the aim of finding out more about their function.
- Identifying entities (e.g., genes, proteins, metabolites) from high-throughput experiments.
- Sequence-based searches, with the aim of finding the identity or the closest relatives of a sequence.
- Submitting data to the public databases, often as part of the publication of a paper.
- Handling of large datasets.
- Statistical analysis.
- Integrative analysis of different ‘omic’ approaches, to gain a systems-based understanding (e.g., combining proteomics, transcriptomics and metabolomics results to find out what’s going on in a particular cell type).

(2) Power users

Power users are typically computational biologists who either aim to understand the biology behind extant data sets, using computational analysis, or whose main focus is supporting end users in their team/department/organisation. Their training needs are distinct from those of end users, often revolving around the need to administrate and integrate diverse biological data sets and gain programmatic access to them. They can be subdivided into two further types of user:
2a: research-oriented power users

These are scientists whose aim is to understand the biology behind extant data sets, using computational analysis. Typical bioinformatics-related tasks may include:

- Collecting data from diverse sources and building bespoke analysis pipelines.
- Writing algorithms to analyse data.
- Data modelling and simulation.
- Downloading test data sets.

2b: service-oriented power users

These are bioinformaticians whose main focus is supporting the experimental biologists in their team/department/organisation. Typical tasks include:

- Downloading data resources, installing and maintaining local copies.
- Combining publicly available data with the lab’s unpublished data.
- Writing algorithms to analyse and/or visualise data.
- Constructing data analysis pipelines.
- Constructing bespoke data resources with a particular research project in mind.

(3) Trainers

Trainers are sufficiently competent users of a data resource that they can train other users. They need to have a firm understanding of users’ needs so that they can tailor their training accordingly and structure their courses to be of high value to a diverse range of users. They also need a firm understanding of different learning styles, so that they can structure their courses to be of high value to a diverse range of users. Trainers are an important stakeholder group for ELIXIR’s training infrastructure because through supporting them to do their job effectively, we will be able to reach a far wider number of end users and power users.

These different types of user are not necessarily mutually exclusive; for example, trainers are often also power users; research-oriented power users often find themselves becoming service-oriented power-users to other groups; and a growing number of researchers are splitting their time between the lab and the computer, spanning the needs of end-user and power user. Nevertheless, it is helpful to keep these different types of audience in mind when designing courses, and to be explicit about which type of audience a particular course is aimed at.

The ELIXIR training infrastructure will have to address the needs of these different audiences, and should ensure that modules and courses made available through the infrastructure describe their target audiences and the goals in sufficient detail to attract an appropriate audience/user base. There are several simple measures that can be taken to ensure this:

- Stating the learning objectives for each course/module (e.g., ‘by the end of this course you will be able to…’)
- Stating any course prerequisites (e.g. should you have done ‘First choose your database’ before launching into a week’s course on one of the databases? Do you need any programming skills? Do you need to be familiar with Unix?)
- Being clear about the limitations of the resource in question (e.g. end-users sometimes think that, if they attend a sequence searching course, we’ll be able to tell them ‘use X and it will do everything for you’; they may be disappointed when we tell them ‘try X, Y and Z, tweaking the parameters for each, to see what will give you the best results for your particular experiment.’ By educating our users
from the very beginning that bioinformatics is an experimental tool, not a panacea, we may be able to avoid such disappointment.

3.2.2 Data resource and tool providers

The data resource and tool providers defined by workpackages 2 and 12 are important stakeholders in developing a bioinformatics user-training infrastructure because there has to be effective communication between database/tool developers and trainers to ensure that user training keeps pace with development of the resources and tools. To achieve this, the ELIXIR training infrastructure will have to incorporate systems for communicating major updates of the resources to those who create the training materials. This is addressed in more detail in section 4.4.

Feedback from training can also provide important input into the further development of data resources and tools. Establishing a two-way dialogue between developers and the trainees, via trainers, could be an important secondary goal of ELIXIR’s user training infrastructure.

3.2.3 Funders of infrastructure

User training is possibly the most effective means of outreach to user groups and potential users. It ensures that funders realise the full benefit of their investment in the development of the data resources and tools by providing users with the power to make the most of these data resources and tools. The case has often been made that the cost of storing biological data and making it available to users is tiny compared with the cost of generating the data in the first place. Similarly, the cost of a training infrastructure for Europe’s core biological data resources and tools may be small compared with the cost of developing and maintaining the data resources. This small investment will make a big difference to the effectiveness of ELIXIR, enabling funders to realise the full benefit of their investment in bioinformatics infrastructure.

3.3 Needs and barriers to progress

As described in more detail in Section 4.4, the need that is specifically addressed by this work package is to provide support and coordination to the various bioinformatics user-training efforts that exist already or will be developed in the future throughout Europe.

So far, this has been hampered by the lack of coordination among the many providers and trainers and the difficulty for trainers in keeping up with the evolution of the resources. Section 4.3 describes this issue in more detail.

A pan-European strategy also needs to take into account that, by its own nature, and because of its rapid evolution, bioinformatics user training requires the involvement of the developers who are responsible for setting up and continuously updating the resources. This information needs to be communicated to the user base, which is enormous and growing. It is important to develop a strategy that helps not only the users, but also the developers, reducing the load on the latter while increasing the benefits to the former. Section 4.4 details our recommendations for achieving this goal.

At the same time, a system for providing an appropriate certification system for trainers, ideally recognised by the scientific and academic communities, would increase the effectiveness of user training throughout Europe, by providing appropriate incentives for trainers.

Finally, several training models can be implemented (see Section 4.1 and Appendix 1), and a pan-European strategy for coordinating their activities should take into account their specific characteristics.
3.4 Challenges and risks

Europe faces several important challenges towards the development of a sustainable bioinformatics training infrastructure. Most importantly, the fragmentation of bioinformatics training activities needs to be overcome, and emerging standards for the development of training modules will have to be consolidated and implemented, in order to promote a structured training approach and a consistently high level of knowledge transfer to users.

ELIXIR plays a pivotal role in the coordination and setup of a structured training environment on a European level. Existing national and local training infrastructures will continue to be the primary producers and providers of new training activities, but ELIXIR needs to integrate these, promote best working practice and benchmarking systems, and provide centralised access (via the development of an online training information portal) to bioinformatics training opportunities across Europe.

ELIXIR also needs to stimulate the development of standard training repositories and learning platforms, which allow for both distributed hosting and development of new activities. A major challenge will be the development of a pan-European training registry. The purpose of such a registry will be to facilitate access to training materials and to enable trainers throughout Europe to contribute to their evolution. The benefits of such a system are clear: sharing the burden of producing and maintaining training materials will reduce duplication of effort, and updates to the data resources and tools will rapidly be reflected in updates to the training materials. The registry will also provide a mechanism for disseminating these updates to many trainers, and will act as a hub for building a community of trainers. The effective exchange of metadata is required to allow linking of training modules from different repositories together, and to register available resources. ELIXIR should therefore ensure the provision of the centralised database and indexing infrastructure, as well as the development and hosting of the training information portal.

There are parallels here with the EU-funded EMBRACE project, which recently built a registry for bioinformatics Web services (www.embraceregistry.net). EMBRACE recognised twin issues: (1) the problems of finding a suitable Web service, of discovering its operational status, and of knowing whether it actually functioned ‘as advertised’; (2) the need to collect and advertise the growing number of databases and tools developed by the project partners, coupled with the need within the consortium to share experiences about the provision and use of Web services. The benefits of having a single point of contact for finding Web services, and for sharing user experiences, were obvious. Various repositories had already been developed for this purpose but, as none was able to determine whether the logic of the service was functioning, it was still commonplace to find services that were broken or no longer maintained.

These issues motivated the development of the Web-service registry. This allows users to register, annotate, monitor and search for services, and acts as a Web2.0-style community server, putting users and providers in touch with one another. Unlike passive mechanisms for recording the existence of Web services, it actively monitors the registration and ongoing behaviour of a service, giving providers and consumers up-to-date status notifications by email or Twitter when a service behaves unexpectedly. The approach embodied in the registry is to recommend a set of industry-standard technologies and to provide tools that help developers move towards adopting these, at the same time recognising that a variety of other approaches exist for pragmatic or historical reasons. The registry therefore allows all manner of services to be added, and aims to provide documentation and support for
users wishing to bring their services in line with standard practices; it hence lowers the barriers to adoption and actively encourages best practice.

The echoes here in terms of the distribution and ‘status’ of training materials and modules across Europe, and the need for an organising registry, are striking. Getting the architecture of such a registry right will be essential from the outset; but we do now have a good model to build on. The EMBRACE registry essentially has three components: a central database, a test harness and a community server. The community server acts both as a readily customisable Web2.0-style foundation (allowing user registration and management, forums, blogs, tagging, rating and search facilities), while the test harness executes the tests for a particular service and reports its status to the database. Building an appropriate registry for training materials and modules will require a similarly principled approach, with input from trainers, trainees and developers, and will help to sow the seeds of the community building that we intend to stimulate further through the repository itself.

While the development of modular training sessions is essential for the custom adaptation of training material to the needs of national and local institutions, ensuring consistent quality and validation of the training material is equally important. National academic institutions should guarantee training and accreditation of the trainers, but accreditation systems need to be internationally recognised and uniformly implemented across Europe. ELIXIR will need to play a central role in stimulating the development of new models for accreditation and validation, and promote international recognition of academic credit, copyrights and licensing of training modules.

The success of this endeavour is closely linked to the capacity of ELIXIR to ensure the coherent integration of both ‘bottom-up’ and ‘top-down’ approaches to bioinformatics user-training provision across Europe. There are several risks associated with such a project:

• The scope and the dimension of the project are difficult to estimate. Bioinformatics data are being generated at an exponential rate, and new tools are constantly being developed to exploit new datasets. Thus, without an appropriately dimensioned infrastructural backbone, ELIXIR may simply not be able to cope with the ever-increasing flood of information and overcome the existing fragmentation of training activities and resources.

• The current bioinformatics training activities are not evenly distributed across Europe, and there is a risk that not all countries will be able to participate at an equal pace in the initial investment to set up new training structures. This could lead to poor uptake of the initiatives in their scientific communities, and lead to further imbalance in the access to bioinformatics resources. ELIXIR thus needs to make specific endeavours to ensure the availability of training opportunities, and stimulate the usage of bioinformatics resources in those countries.

• Lack of sustainable funding methods. The success of this project will not only depend on the design and setup of the initial training infrastructures across Europe, but also require long term funding to maintain the central and human resources to guarantee the coordination and management on a European level in a sustainable manner.

• Without proper stakeholder engagement, there is a risk that ELIXIR might be considered to be imposing a structure, leading to poor uptake of the training resources that we discuss later in section 4.4.1. ELIXIR’s training infrastructure must therefore make every effort to build a community of training stakeholders from the very beginning, and steer its stakeholders away from a purely top-down approach, which would stifle creativity in training and reduce ELIXIR’s impact.
Given the essential importance of bioinformatics as a core component of technological development in the life sciences, and its potential impact on modern medicine, agriculture and environmental science, these risks are easily outweighed by the negative impact of not undertaking this project. The general efficiency of non-structured, individual training approaches is much lower than the integrated approach proposed by ELIXIR. Maintenance, compatibility and security of resources is hard to ensure, and in the long term would be wasting many more resources than those required by ELIXIR for the initial setup of a structured European training landscape.

4. Description of roadmap in terms of elements/key themes

4.1. State of the art

4.1.1. Current training methods and technologies

The focus of effective bioinformatics user training is on helping users to get the best out of a data resource or tool, to understand which tool is appropriate for which task, and to understand the relative strengths and weaknesses of related tools and resources. We define various means by which this may be achieved:

Face-to-face training

This is the gold standard of user training. Typically, users are trained in small groups (10–30), either by expert users of the resources or by people whose job it is to maintain or develop the resources. Ideally, time is built into the programme to give the trainees quality networking time with the trainers, so that they can ask questions and relate what they’re learning to their own research problems. There are several variations on the theme of face-to-face training, each with their own pros and cons:

(1) Users travel to a centre of excellence. This model is widely adopted for bioinformatics user training, and for software training in general. Examples include training programmes at the The Gulbenkian Institute, the University of Cambridge and the Centro de Investigaciones Principe Felipe, Valencia, Spain (see Appendix 1).

Strengths of these programmes include:

- Total immersion in the subject area for the duration of the course;
- Quality contact with the trainers;
- Access to other local experts if a trainee has a question that is beyond the area of expertise of the trainer;
- Quality-controlled training materials and well-equipped training rooms.

The weaknesses are:

- the expense of sending trainees on a residential course;
- lack of scope to modify the programme to the needs of individual trainees or small groups of trainees from the same institute;
- the intensive nature of these courses, and their often rapid pace, makes them unsuitable for individuals who prefer to learn in their own time.
- The tendency for local trainers to ‘drop in’ for their training slots but not to engage with the course as a whole; without careful coordination this can lead to redundancy in the course materials, inconsistencies in training methodology, and insufficient networking time with the trainees.
(2) Trainer(s) come to trainees (‘roadshows’)

This is another widely adopted system, both for bioinformatics user training and for software training in general. The Bioinformatics Roadshow that formed part of the EU-funded FELICS Integrated Infrastructure Initiative and is being continued through the SLING Integrating Action is a typical example (see Appendix 1); individual data resources also regularly provide training in this way. This type of training programme offers a great deal of flexibility. Training can be tailored to the facilities available and to a more homogeneous audience with shared interests. Strengths of roadshow programmes include:

- flexibility for the host – the programme can be tailored to the needs of the host institute;
- low cost: it’s cheaper to fly 2-4 trainers to the trainees than it is to fly upwards of 10 trainees to a dedicated training facility.

Weaknesses include:

- lack of control over the training facilities: software or networking problems are more frequent than they are in courses run at dedicated training facilities and this can have a significant impact on the value of the training. The host’s concept of an appropriate training facility may not match up to the trainer’s assumptions.
- fewer opportunities for networking: roadshows tend to be larger and less personal than courses held at centres of excellence, with the trainees dipping in and out of sessions to fit them around their normal working day. This reduces interactions with the trainers and removes the opportunity for the development of collaborations with other trainees.
- Lack of immediate access to experts other than the trainer.

(3) Summer-school-style courses

Another widely used format is the summer-school format, in which both the trainers and the trainees travel to a neutral venue. The BioSapiens Network of Excellence, ICGEB, Wellcome Trust Advanced Courses and the University of Adam Mickiewicz in Poznan run courses of this nature. This has several advantages: the trainers, being away from home, are more likely to immerse themselves in the course, providing the best possible opportunities for interaction with the trainees, and a convivial atmosphere that is extremely conducive to learning and to the development of long-standing collaborations with other scientists. The disadvantage is the added cost incurred by trainers and trainees travelling.

The BioSapiens European Bioinformatics School partially addressed this issue by rotating from country to country, thereby spreading access to training as widely as possible. The budget of the Network of Excellence provided funds to reduce the costs for trainees, and also involved the post-docs hired by the Network in the teaching. Both aspects turned out to be instrumental in making the school available to participants from all over the continent. ELIXIR may be able to build upon this success by providing support for trainers who wish to adopt a similar approach (see section 4.4.1).

E-learning

The advantages of e-learning are obvious: it has the potential to reach a much wider audience than face-to-face courses can, and therefore has the potential to be more cost effective. It is also independent of geography and, when adequately funded, can be made freely available, making it accessible to trainees who do not have the funds to travel. We divide e-learning into two categories:
(1) trainer-led e-learning

At their simplest, e-learning courses can be created by capturing the activities of the trainer in real time and broadcasting them, either live or as prerecorded courses. The CSC in Finland adopts this methodology, as do many commercial software providers. There is an art to capturing lectures and demos in this way; short videos are preferable (the trainee’s attention tends to drift when there is no eye contact with the trainer) and demonstrating complex websites at the low-resolution used by most webcasts is a significant challenge. Lectures and demos can easily be archived and shared for future use, but have a limited lifespan for rapidly evolving bioinformatics resources.

Interaction with the trainer, and with other trainees, enriches any learning experience. The Center for Biological Sequence Analysis (CBS) in Denmark has developed a method that combines classical face-to-face training with e-learning. The method aims at offering courses to the largest possible audience while at the same time keeping the added cost and effort of e-learning to a minimum. Here, training sessions are simply broadcast versions of the classroom teaching with the addition of online students.

Strengths of trainer-led e-learning include the following:

• It combines the benefits of both face-to-face training and trainee-led e-learning (see below)
• It is easy for the trainer as well as the students; only the chat moderator (or a third person) needs to be familiar with the technical aspects of the system.
• It is cost effective, as classroom-style training can be provided without the trainees having to travel
• It provides a livelier learning environment than pre-recorded e-learning courses, whilst providing the option to record the sessions for later use (although clearly not in any real interactive form).

Weaknesses include the following:

• Online students generally have no direct interaction with the teacher or other students
• It is live, which means that things can go wrong (loss of sound, video feed, or other technical issues).
• A reasonable bandwidth is needed both on the trainer and on the user side to allow a smooth experience; this can limit accessibility to those regions that may benefit the most from e-learning.
• Reuse of material is limited to recordings.

(2) Trainee-led e-learning

The second category of e-learning resource comprises standalone courses that can be taken in the trainee’s own time. Several such resources have been developed, including the European Multimedia Bioinformatics Educational Resource (EMBER; an EU funded initiative led by the University of Manchester, with ten partners); eProxemis (developed by GeneBio) and the EBI’s pilot e-learning portal.

Although all of these were designed to provide a self-contained, interactive learning experience, they have found other uses in a variety of different contexts:

• As the practical component of traditional face-to-face training courses (including bioinformatics summer schools, CPD programmes, etc.) and Master’s modules;
• As self-contained modules on distance learning Master’s programmes (for example, EMBER has been integrated into the distance learning Master’s programmes at Manchester and Birkbeck College) and Graduate Training Programmes; and

• As self-contained, standalone courses on the Web, freely accessible by the community, for individual self-paced study or inclusion within courses worldwide.

Standalone courses tend to use a variety of different teaching methodologies, catering to different learning preferences. These include tutorials and demos, quizzes, practice sessions, and occasionally forums and chat rooms that facilitate interaction with experts and/or other trainees.

Several different e-learning frameworks have been used – some purpose-built for the project (EMBER, eProxemis), others using off-the-shelf products (the EBI’s e-learning portal uses moodle, whereas EMBER is now being ported to Blackboard). Strengths of trainee-led e-learning courses include:

• Trainees can learn in their own time and at their own pace;

• The courses complement face-to-face training, providing a means for trainees to build on what they’ve learned in class.

Weaknesses include:

• The extremely short time it takes for information to go out of date;

• The high costs (both in terms of time and in terms of money) of developing material (especially videos) and ensuring that courses are internally consistent: updating one aspect of a course (for example a tutorial) has knock-on effects for other parts (e.g. videos, quizzes and reflective tasks)

• The lack of instant availability of a trainer or other expert to ask;

• The lack of defined start and end times: this type of training is best suited to highly self-motivated individuals who have the willpower to fit learning around their day-to-day activities.

4.1.2 Survey of current training availability throughout Europe

To gain a more comprehensive overview of how much bioinformatics training is ongoing in Europe, the ELIXIR training committee conducted a simple survey. Through mailing lists, we asked previous attendees, course organisers and general contacts to send us information on available bioinformatics courses. We specifically gathered information for those courses open to all users and excluded those belonging to closed educational programmes. We used a form containing predefined fields and drop-down menus to gather, in a standardised and comparable format, the information about available bioinformatics training in Europe. We focused on courses that are re-run on a regular basis, as opposed to one-off courses, because we wanted to understand which courses are run regularly so that we can identify any gaps.

The data presented in this report contains information that we gathered from 8th May 2008 until 18th September 2008. This survey has revealed that the vast majority of the training (62 courses were used for the analysis; we now have information on more than 80 courses) reported on takes place in Western Europe, with the overriding majority of regularly repeated user-training courses taking place in the UK (Figure 1).
Figure 1. Geographical distribution of courses as of November 2008.
A clickable map of the courses featured in the survey is available at http://www.elixir-europe.org/page.php?page=user_training_map (Figure 2).

Figure 2. Detail of the ELIXIR training map, showing regularly repeated bioinformatics user-training courses available throughout Europe.
The most frequently covered topics are transcriptomics, general overviews of bioinformatics, and programmatic access to data resources. These three topics account for almost half of the training documented in the survey: (Figure 3).

Figure 3. Breakdown by topic of bioinformatics user-training courses currently run in Europe.

When we consider the individual data resources covered, there was a wide distribution of different resources, with most of Europe’s core data resources featuring in the top 10: (Figure 4).

Figure 4. Breakdown by data resource of bioinformatics user-training courses currently run in Europe.
The target audience for most of the courses included in our survey was end users: (Figure 5).

![Figure 5. Breakdown by target audience of bioinformatics user-training courses currently run in Europe.](image)

While a survey of this nature cannot hope to be complete, the following conclusions can be drawn from it:

- In Western Europe, user-training courses are being run on a regular basis covering a wide range of data resources and tools, and including Europe’s core data resources. These courses are mostly aimed at end-users (or at least are perceived to be aimed at end users by the people who completed the survey).

- In the new EU member states, EITHER little or no user training is taking place, OR the survey failed to reach trainers and course organisers in these countries.

- Although there are regularly run courses that cover many of Europe’s core data resources and tools, provision is heavily skewed towards a few topics. This may reflect demand, but it will be important to determine whether this is the case so that we can best address the training needs of users in the future.

### 4.1.3 Surveying different training methodologies

ELIXIR’s training work package also undertook a technical feasibility study to gain insight into whether e-learning can be an effective vehicle for bioinformatics user training. This study used the same training materials, presented to audiences with similar academic backgrounds, in both a face-to-face training course and a trainee-led e-learning environment.

For the Face-to-face ‘arm’ of the study, the second day of the 8th BioSapiens School of Bioinformatics, held in May 2008, was devoted to a sequence searching course. The trainers used materials that had been developed for the EBI’s pilot e-learning project to teach the 40 course delegates, using a combination of lectures (based on tutorials developed for the sequence searching e-learning course) and hands-on training (using quizzes and reflective tasks developed for the e-learning course).
For the e-learning arm of the study, volunteers were invited by e-mail and on the EMBL-European Bioinformatics Institute’s website to sign up to beta test the pilot e-learning course on Sequence Searching.

The two groups were compared using a 10-question quiz designed to test their understanding of the course material. Although the numbers are small and we question their statistical significance, we found that the mean quiz score of both groups was similar, but the average time taken for the e-learning arm to complete the quiz was shorter (Figure 6).

![Face-to-face vs e-learning](image)

**Figure 6.** Distribution of quiz scores versus time taken to complete a course in two cohorts of students, one taught in a face-to-face learning environment, the other taught the same course in an e-learning environment.

We also found that the spread of scores was larger for the e-learning group, with a slightly higher fail rate (Figure 7).

![Marks and Failures](image)

**Figure 7.** Distribution of marks and numbers of failures in a short quiz taken after face-to-face training versus the same course taught in an e-learning environment.
Whilst much larger numbers would be required for us to gain conclusive information from this study, our initial impression is that trainee-led e-learning can be an effective means of doing bioinformatics user training, and that this methodology should therefore be included in ELIXIR’s training strategy.

4.2. Specific opportunities

Bioinformatics has become a core element of pure and applied life-science research in the 21st century. ELIXIR therefore provides a new set of opportunities for the development of the European Research Area through the design of a new approach towards bioinformatics user training that may be applicable more widely:

• Bioinformatics datasets are the most well structured and universally shared scientific data resources currently available. The existing structure in the underlying data is, however, not yet reflected in the availability of similarly structured and shared training resources. The development of a pan-European training infrastructure is therefore not only a challenge in itself, but has the best potential to serve as a pilot project to generate the necessary expertise for the implementation of shared international training resources in general. This expertise may be employed in other areas of life science, as well contributing towards advanced training opportunities in other areas of tomorrow’s knowledge-based society.

• Bioinformatics software tools are frequently developed with a specific goal in mind. Opening up their availability and accessibility through collaborative training efforts not only greatly benefits the international research community, but also provides important feedback to software developers, enabling them to further improve their resources. Developers will also be stimulated to participate in training development, which in turn will help them to adapt software to users’ needs, provide better documentation, and thus greatly increase the return on investment in research funding on a European scale.

• The development of a central training information portal, with associated ‘user forum’ capabilities, will enable bioinformatics user communities to share information in a much better way than is currently possible. It will also stimulate the development of new tools based on user demand, through the direct interaction of tool developers and end-users. By providing a general overview of existing software and training resources, the central portal will thus generate more streamlined and modular software, and will facilitate the maintenance of up-to-date documentation and training material. This in turn will make user training more effective and beneficial for the research community.

• Bioinformatics training is not yet part of standard life-science curricula in all European countries. The availability of standardised and modular training materials will facilitate development and implementation of specifically tailored courses in countries that have not yet developed a sufficiently large user base to warrant the development of their own local training opportunities.

• The overview of training-module availability through a central information repository will greatly enhance the development and deployment of tailor-made courses, which should also stimulate the further integration of bioinformatics into standard university, and possibly even high-school, curricula, in turn promoting interest in information technology and scientific careers early on.

• Bioinformatics training requires the use of computers and employs internet-based informatics resources. These characteristics make it particularly suitable as a target for new web-based and e-learning training methodologies. The
development of new software tools to facilitate and standardise the deployment of
e-learning and web-based training modules to the end user would benefit not only
the bioinformatics community, but also general scientific training efforts in other
areas of research, and possibly feed into educational training projects as well.

These are only a few of the specific opportunities associated with this ELIXIR’s
training infrastructure. Structured training is a central pillar of scientific and
technological progress in the European Research Area, and the importance of further
investments into this area cannot be emphasised sufficiently.

4.3. Specific barriers to progress
While there are many opportunities, there are also a few potential barriers to
progress. These barriers are not necessarily specific to this project, but are a more
general problem for promoting access to biological research infrastructures on a
European scale. Most of those barriers are also challenges and opportunities that
ELIXIR is specifically designed to master, thereby improving advanced training
opportunities in Europe.

ELIXIR will pave the way towards the establishment of shared pan-European training
resources. However, there is no current ‘European way’ of training bioinformatics
users. In the multi-national context of the European Research Area, this might be a
barrier, but at the same time should be considered an opportunity to bring research
communities together and build on the long-lasting efforts of the European
Commission to improve international mobility aspect of researchers across Europe.

Bioinformatics datasets and software tools are in constant flow, and change rapidly
and steadily to keep pace with new developments in many scientific areas. This
might cause difficulties in keeping training resources up to date. However, a
centralised repository of training materials that are benchmarked by their users
allows a much larger user community to share information, and encourages end
users to directly communicate with developers and trainers, thus ensuring a fast
turnaround of documentation and training resources.

Bioinformatics training frequently does not receive academic accreditation and thus
training is only provided in an ‘ad hoc’ and unstructured manner. ELIXIR will be able
to promote the accreditation of bioinformatics training activities by national academic
institutions, thus increasing the acceptance of bioinformatics as part of standard
curricula and promoting information technology as a career path in the life sciences.
The current lack of recognition that user training makes an important and valuable
contribution to European research will have to be addressed if we are to motivate
highly qualified and busy scientists to dedicate time to training.

Bioinformatics training activities are not equally distributed across all countries, and
lack of sustainable funding in some countries may prohibit participation at an equal
pace in the initial setup of new training structures. ELIXIR thus needs not only to
ensure central core funding, but also to promote local training activities to receive
sustainable funds to ensure their visibility and acceptance by the local scientific
community.

4.4. Specific challenges to be addressed

4.4.1. Specific activities
We envisage ELIXIR’s training infrastructure to comprise two parts (see Figure 8).
Figure 8. ELIXIR’s recommendations for a pan-European training infrastructure. The infrastructure (top lozenge, orange) comprises two parts: (1) a mechanism by which the development of data resources, tools and standards is tightly linked to the provision of training materials; (2) a training support unit that will: (a) provide a centralised training registry that will improve access to bioinformatics user training throughout Europe; (b) provide support for trainers throughout Europe, so that they can use the materials collected by ELIXIR to develop training courses that are directly relevant to their local trainees; (c) develop benchmarking and evaluation systems; (d) provide mechanisms for developing, piloting and evaluating new training programmes; and (e) act as a single point of contact for national and pan-European training infrastructures and projects. Other infrastructures with which close coordination will be vital (grey lozenge, bottom) include the ESFRI Biomedical Science Infrastructures (green rectangles) and the Innovative Medicines Initiative Education and Training call topics (orange rectangles).

4.4.2 Training as an integral part of service provision
The first part relates to developing a set of principles by which ELIXIR’s data resources (WP02), tools (WP12) and standards (WP07) operate. Principles such as making data and source code freely available, and compliance with community standards, are well established in the bioinformatics community and are likely to be prerequisites for all ELIXIR-funded bioinformatics resources (see Workpackage 2’s report). To this we propose adding some minimal requirements that will facilitate the creation and maintenance of training materials. The creation of these standards will need to be an early action item for ELIXIR’s implementation phase, and we do not provide a comprehensive list of such standards here; we anticipate that they would include information on:
• The purpose of the resource/tool/standard
• Intended users
• Periodicity of update cycles
• Helpdesk/bug reporting contacts
• Advanced warning of upcoming new features and major updates

Some of this information could be built into a series of fact sheets, akin to EMBnet’s quick guides (www.embnet.org/QuickGuides) or the EBI’s resource fact sheets (www.ebi.ac.uk/Information/Brochures/). Both have proved immensely popular with users and are an effective way of disseminating structured information about biological data resources.

To fulfil these requirements, we propose that provision of adequate training/outreach staff be built into the funding mechanism for every ELIXIR-funded data resource/tool/standard. Perhaps the simplest way to do this would be to dedicate a certain percentage of the headcount for each data resource to documentation, outreach, training and supporting other trainers. Various models could be built for different types of tool/resource/standard based on successful outreach and training programmes for existing bioinformatics services. These individuals would be responsible for providing content for the ELIXIR user-training registry (see section 4.43).

One factor that we see as crucial to supporting effective training is the provision of use cases that are directly relevant to the user community and that remain robust with each new release cycle. We propose that the testing of such use cases be built into the development cycle of each resource/tool/standard so that it would be guaranteed to work after each release. This is one reason why we feel it necessary to place outreach and training personnel within bioinformatics service teams, because this improves the chances of effective communication between developers. Another advantage of this is that feedback from trainees can readily be given directly to the developers, facilitating the creation of user-friendly resources.

All too often, trainers spend a lot of time developing tutorials only to find that a significant change to the data resource or tool renders their tutorial materials obsolete. We therefore propose that an ‘early-warning system’ be built into the development cycle of ELIXIR’s data resources and tools to alert trainers and training coordinators to any major changes that might have an impact on user training. One way of doing this might be to include an ‘upcoming new features’ alert in the prerequisites for ELIXIR data resources and tools.

4.4.3 The ELIXIR training support unit

One efficient and cost-effective solution to the geographical inequalities in user training would be to create a mechanism for supporting ‘local trainers’ to train the scientific and medical communities within their own organisations or countries. This does not mean dictating how they should approach their training, nor passing judgement on the quality of their training practices. On the contrary, it is about finding the most appropriate strategy to support and facilitate their work across Europe.

There are two distinct strands here: one would simply involve collating, centralising and disseminating information relating to training that is ongoing, or planned; the second would be somewhat akin to the notion of Continuing Professional Development (CPD), but with a community rather than an individual focus. Both functions require a level of coordination that would best be addressed through a dedicated unit tasked with supporting bioinformatics user training throughout Europe. We see this unit as performing the following functions (see Figure 8):
(1) Providing a centralised training registry

The idea here would be to develop and maintain a central repository of, or portal for, information relating both to ongoing and to future training activities. Clearly, different types of information will be important or relevant to different users, so any such repository would need to be comprehensive and carefully structured so that it could both serve, and be informed by, a diverse user base. The kind of content to which such a portal might provide access would include information on:

- Existing training modules (what they are, what they include, where they happen, their duration, etc.);
- Planned training activities (this could include courses that trainers are preparing, both to give advanced warning to potential trainees, and to solicit input in terms of additional content and hands-on help from other trainers);
- Training materials that are freely available either to download from the Internet or to access via the Web (these might include video lectures, animations, Powerpoint presentations, interactive tutorials, worked examples, data suitable for developing new exercises, and so on).
- A database of trainers, so that course organisers can find appropriate experts.
- A directory of:
  - Institutions that offer training opportunities;
  - Literature that supports training activities;
  - Organisations or groups both inside and outside Europe that offer education and training content (the ISCB, EMBnet, APBioNet, Bioinformatics.org, etc.);
  - Funding bodies that offer, for example, travel fellowships for pursuit of training activities.

To be maximally useful and to best serve the life-science research community, the community itself would need to involve itself in contributing the most up-to-date information to the portal. Early efforts to engage with this community will therefore be an important element of building the portal.

To map this back to our stakeholders, bioinformatics service providers and trainers would submit training materials and associated metadata that would allow the material to be categorised. Trainers throughout Europe would be able to search for training materials and find the most appropriate ones for incorporation into their own training courses. Users (trainees) would have ready access to information on upcoming training activities.

The training-support unit would not need to perform an editing function, and need impose only minimal guidelines on the format of the training material. The rest would be up to the submitter. ELIXIR could implement a standard metadata format for describing training material. This could be very simple at first, but extendable if needed. The metadata would be an integral part of any training material submitted to the portal. This will enable the automated sharing of training materials across multiple sites through the central repository, and would facilitate the dispersed storage of training materials. Parallels with existing e-learning standards, such as SCORM, are notable here. When training materials (or links to it) are submitted to the central repository, the receiving system must at least semi-automatically generate metadata describing the material.

This system would support multiple teaching styles, ranging from traditional (and still extremely common) teacher-centred approaches, where the activities are given by the teacher, and trainees have little or no control of their goals, to more recently
developed approaches that give some or even most of the control over activities to the trainees. The portal could help ELIXIR to do this because its focus would be on common training materials, not on common training courses. The challenge will be to develop a system that allows training materials to be categorised in such a way that trainers can identify relevant materials for adaptation to fit the curricula of their own organisations. The development of a system that asks submitters to distinguish between “raw” materials and those that have been developed for a particular target audience or a particular learning preference may go some way towards this.

Any material submitted would need to be openly accessible to, and modifiable by, other trainers. One way of achieving this would be to request that all submitted materials be made available under a creative commons share-alike/attribute or similar license (http://creativecommons.org/about/licenses/) so that others can adapt it without restriction, whilst ensuring, as far as possible, that the original creator was acknowledged and the information retained in the public domain.

This system relies on bioinformatics experts dedicating a significant amount of time to developing and categorising training materials. To ensure that performing this role is given the credit that it deserves, it will be important that authors are given appropriate recognition by the research community for supplying material. This could be based on existing national models (for example, the Finnish Virtual University: http://www.virtuaaliyliopisto.fi) in which validated material is given the same status as a publication. It may also be possible to give ISBN numbers for validated materials.

Validation and updating of training material must be straightforward for authors and co-authors. Updating must also be simple enough for teachers to be motivated to do it. Validation does not necessarily mean a process where the information in the material is approved, but rather an administrative process whereby overlaps with other older materials provided by the same authors are removed, or where links to related material (for example, derivative materials produced by others) are created.

A critical success factor will be the motivation of teachers to use the materials provided through the ELIXIR training infrastructure. Ease of use and availability will be keys to success once the service has been established and effectively communicated to target users. National and international networks of trainers will be the key channels.

Building effective channels of communication with the user community will be an important aspect of this activity, so it will be necessary to create an outreach plan to users, trainers and training coordinators, and to set up appropriate channels of communication, such as mailing lists, FAQs, surveys and social networking resources. Surveying the training needs of the user community will enable ELIXIR to respond to emerging research trends.

One of the greatest challenges to users in the current bioinformatics training climate is finding out which user training courses are available and what the differences are between them. As there is currently no standardised way of describing course content, it is impossible for users to compare courses, and significant overlap between courses may therefore be obscured. We propose that ELIXIR develops a simple calendar-based system, compliant with commonly used standards, such as ICS or CalDAV, to gather information on bioinformatics user-training courses taking place throughout Europe. We have already made some progress with this as part of a data-gathering exercise (see section 4.1.2). In so doing, we have attempted to break courses down into training modules, on the basis of the data resources and tools covered in each course. This has proved to be a challenge because it is often very difficult to ascertain information on both the subject areas and the depth of information covered simply from a course programme. We propose to request structured summaries for each course that will allow potential trainees to identify:
• Data resources/tools covered
• Target audience
• Learning objectives
• Course prerequisites
• Supporting publications
• Availability of travel bursaries

As well as facilitating access to training throughout Europe, the calendar will allow ELIXIR to identify gaps, both in terms of geography and in terms of resources covered. This will help to identify priorities for the development of new training materials and courses.

We favour this light-handed approach towards creating a resource for supporting trainers because it is less susceptible to bottlenecks than a more tightly quality-controlled system. Furthermore, it is inevitable that individual institutions will want to develop their own training resources and programmes, each one having its own preferred format. The above system allows this without tying anyone to a specific format or platform. Crucial to the success of such a system will be a means of benchmarking the content submitted to the registry (see point 2 below).

(2) Developing benchmarking and evaluation systems

A benchmarking system will allow users to identify relevant training materials in the ELIXIR training registry. The simplest system might be to allow users themselves to rate the training materials and add reviews, similar to the book reviews on amazon.com.

A related, but distinct, function is to support trainers to evaluate their courses. In our experience, every trainer has his/her own set of benchmarks by which they evaluate their courses, so we feel that it would be misguided for the ELIXIR training support unit to impose a single evaluation system. A pragmatic approach would be to enable the sharing of best practice by collecting examples of successful evaluation systems and tools, and collating a core set of feedback questions that could be incorporated into all course evaluations to aid comparison of different training approaches.

(3) Providing (or coordinating) support for trainers

The idea here would be to build on the concept of CPD, initially aiming to provide opportunities for individuals to learn about a range of specialised methods and technologies in order to maintain, develop and enhance their own skills and knowledge. As noted above, the idea goes much further than this, as the motivation is not just to enhance the skills of individuals, but rather, to help them to take those skills (and the course materials) back to institutions in their own countries, in order to disseminate such specialised training to much broader audiences. As already mentioned, this is important, because it is clearly much more cost-effective (and, indeed, more practicable) to support trainers than it is to try to deliver courses for all European trainees from a single institution or network of institutions. By taking on a trainer-support role, ELIXIR’s training infrastructure would also distinguish itself clearly from educational establishments, thereby reducing the risk that it might be seen as competing with already established educational programmes.

This kind of training model has been exploited very effectively in past European projects and networks, most notably within EMBnet. EMBnet embraces a vast community, from the most to the least affluent countries within (and outwith) Europe. Since its inception (20 years ago), EMBnet has maintained a register of teachers and trainers, of training opportunities, of teaching and training resources, of funding opportunities, and so on, all of which is overseen by a formal Education and Training
Committee. This commitment to training led to the development of the European Bioinformatics Educational Resource (EMBER; see Appendix 1).

These experiences have highlighted the importance of being able not only to expose representatives of institutes in different European countries to state-of-the-art tools and technologies, but also to facilitate their efforts to train colleagues and students at home to the same level.

(4) Enabling the development and integration of training programmes
There are several different ways in which ELIXIR could catalyse the transformation of demand into new training activities:

• ELIXIR funds a programme that allows potential course coordinators to observe a course in action. This will help those who want to set up their own local training programmes to learn from existing, successful training initiatives.

• ELIXIR funds networking sessions for trainers, thereby facilitating the exchange of expertise and innovation among Europe’s community of bioinformatics trainers.

• ELIXIR performs a ‘barometer’ function, communicating changing training needs to funding bodies who could then decide whether funding such training courses was a priority for them.

• ELIXIR has funds set aside to develop and pilot new training activities, and manages an application process for such funds.

• ELIXIR develops and pilots training courses itself; this could perhaps be an evolutionary progression of the BioSapiens European School.

Decisions on the best way forward will require further discussion among funders (including current funders of bioinformatics training courses, such as EMBO, FEBS, the Wellcome Trust and ESF), Workpackage 11 representatives and the ELIXIR Scientific Advisory Board.

(5) Interaction with national and pan-European training infrastructures
ELIXIR sits within a complex landscape that involves other ESFRI biomedical science and environmental science infrastructures, the Innovative Medicines Initiative, and other EC directorates, such as DG-Information Society. It is vital that ELIXIR maintains regular contact with these other bodies for all of them to achieve success. Bioinformatics training is an important part of many of the ESFRI BMS projects, as well as IMI. A central contact point that can funnel relevant information from these other projects to ELIXIR’s stakeholders, and from ELIXIR to the other projects, will be an asset to Europe’s competitiveness in biomedical science.

As well as the infrastructures, there are already a number of projects that aim to make bioinformatics user training available to European researchers. These include: open training courses at centres of excellence, such as the Gulbenkian Institute and EMBL; ‘roving’ training programmes, such as the FELICS/SLING Bioinformatics Roadshow and the BioSapiens European School; and distance Learning programmes such as Oxford University’s Computational Biology masters’ programme, Manchester University’s Bioinformatics Masters and open access Bioinformatics VLE, and EMBER. ELIXIR should neither try to emulate such programmes nor try to control what they are doing. Rather, it should aim to support them by building networks of trainers and contacts with user communities. With the support of ELIXIR’s communities workpackage, the ELIXIR training support unit will be able to build appropriate contacts in countries where there are currently no regular user training programmes, and put them in touch with appropriate training projects.
4.4.2. Implications for funding actors

The implications of ELIXIR’s training strategy for funding actors fall into three categories:

1. A mechanism for funding each of ELIXIR’s data resources and tools to support the development of training materials, as described in section 4.4.2 above. Under the current model for ELIXIR, this would best be done by integrating such funding into the funding model for the resources themselves rather than by ‘bolting on’ a separate training fund. This suggested funding model mirrors our philosophy of keeping database development and training materials development close together.

2. A mechanism for funding the ELIXIR training support unit. Again, it would be most appropriate to fund this at a pan-European level because the unit is specifically designed to integrate training activities, and improve accessibility to training, throughout Europe. Budgeting for this unit will need to take into account other pan-European training projects (e.g., IMI call 14), and we will need to interact closely with these to avoid redundancy and remain cost effective.

3. On a national level, funders will be able to do much to encourage full use of ELIXIR’s training infrastructure. There is a trend towards funding bodies mandating their beneficiaries to submit data to public domain resources, and this has to be backed up by training that enables researchers to do this. For example, the UK’s Natural Environment Research Council devised its own training programme to enable NERC-funded scientists to submit microarray data to ArrayExpress. The sharing of training expertise throughout Europe may reduce the need for funders to develop their own bespoke training solutions. Involving funders in the gap analysis, and taking account of training needs at the national level, will help ELIXIR to develop a cost-effective training infrastructure.

4.5. Priorities mapped to time

4.5.1. Short term priorities (2011-2013)

Our short-term priorities revolve around building a more accurate picture of current user-training availability throughout Europe, and on developing systems for categorising it and performing gap analysis. Specific activities include the following:

• Our first priority is to generate a more complete picture of the training activities that are already available, using the tools that we are already piloting (see section 4.1.2) and input from the national representatives in workpackage 3bc.

• This will enable us to perform a more accurate gap analysis to identify which training modules are under-represented (i.e., for which there is a mismatch between demand and availability) and which countries are under-served by training.

• Both of these activities, and many of the other early-stage activities, require us to build a community of bioinformatics training stakeholders, with representatives from all the stakeholder groups described in section 3.2.

• In parallel, we will develop a controlled vocabulary for standardising course descriptions. This will not only include course contents, but also prerequisites and learning objectives. We will need to interact with other efforts to provide standardised course descriptions (for example, EUFEPS, www.etplatform.eu). This implies a rather substantial effort with dedicated resources in the initial phase.

• We will develop tools for automatically collating information on available training. These tools must make it simple for course organisers to submit information on
their training activities, and equally simple for users to find out what is available and what is relevant to them. Tools already in development include a meta-calendar that collects the information from several decentralised calendars, and a linked Google-Earth map of available training activities.

- We will work with developers of ELIXIR’s data resources and tools to develop recommendations for informing the trainers about updates and changes.
- We will survey and evaluate new technologies for training and provide recommendations about their suitability for different training purposes; our links with IMI call 14 will prove especially useful here.

4.5.2 Medium term priorities (2013-2015)

In the medium term, we will move beyond what is currently available to engage emerging user groups, and will implement our two-part strategy for supporting user training:

- We will identify emerging user groups and work with them to define the types of training required; this will be done in collaboration with other ESFRI BMSs, LifeWatch, relevant IMI call topics, and other pan-European groups with a need to use biological data.
- We will develop a fully fledged training-support portal containing structured information on user training at a pan-European level, taking advantage of the abovementioned controlled vocabulary and of the automatic system for collating available resources.
- We will implement an automatic mechanism for warning trainers about updates to ELIXIR resources.
- We will foster the implementation of a set of use cases that are relevant to different user groups and that are regularly updated to take into account updates to the data resources and tools. This will be done in consultation with the user groups identified in point 1 above.
- We will set up a technology watch to survey and evaluate new technologies for training and provide recommendations about their suitability for different training purposes. Again, there are strong parallels with IMI call 14 here and we will work closely with them on this deliverable.
- We will build in feedback mechanisms from developers, trainers and users to ensure that our methodologies are constantly improved and remain relevant to their needs. This final point will help ELIXIR’s training support to develop in the long term.

4.5.3 Long term priorities (2016+)

At the present time it is extremely difficult to predict the shape of the ELIXIR training infrastructure beyond 2016; this will depend enormously on how technologies for sharing information develop, and how training methodologies evolve. By remaining engaged with our different stakeholder groups and regularly surveying their needs, we will ensure that ELIXIR’s training infrastructure continually evolves to meet the needs of Europe’s life science research community.

What is clearer is that by 2016 we will need to have in place the tools to ensure that bioinformatics user training resources are readily available to any European scientist who requires them, and that mechanisms are in place to ensure that such resources remain up to date, both in terms of content and in terms of teaching methodologies.
4.6. Specific dependencies

ELIXIR will reshape the way bioinformatics data resources are organised in Europe. The role of the training workpackage is to make sure that the benefits of this huge effort are spread throughout ELIXIR’s large and diverse user community. It follows that the training workpackage depends upon interactions with many of the other workpackages, and with other European initiatives, both within and outwith the ESFRI programme. Within ELIXIR, WP2 (Data Resources Committee), WP3bc (Bioinformatics Communities Committee), WP7 (Data Integration & Interoperability) and WP12 (Infrastructure for Tools Integration Committee) are the primary interfaces with this workpackage. As ELIXIR’s training infrastructure is concerned with ensuring that new user groups receive adequate user training, interactions with WP9 (Interdisciplinary Interactions between biological information and Medical/Health and Nutrition Information) and WP10 (Interdisciplinary Interactions with Chemical, Plant, Environmental & Agriculture Databases Committee) are also of strategic importance.

A key aspect of ELIXIR’s training strategy is that it will not only benefit the ELIXIR Stakeholders, but it will be an asset for all other ESFRI Biomedical and Life Sciences projects. ELIXIR’s training committee is already working closely with these other infrastructures, and with EFPIA, through IMI call 14, and there are many synergies with this call topic. Careful coordination with IMI call 14 will enable ELIXIR to engage readily with new user groups in medicines development. Similarly, there are synergies with the agricultural and food industries and with environmental research (e.g., LifeWatch).

ELIXIR’s training strategy will also build upon the training experience of other EU projects, for example the very successful Permanent European Bioinformatics School of BioSapiens, which developed a very fruitful collaboration with the ENFIN NoE for training. The FELICS Integrated Infrastructure Initiative and the SLING integrating action have also done much to bring bioinformatics user training to many new users throughout and beyond Europe, and will help to nucleate ELIXIR’s training stakeholder community. Other NoEs, such as EMBRACE and Gene2Phen, have training initiatives for their own specific user groups, and it will be important for ELIXIR to build upon these.

Finally, ELIXIR’s training infrastructure should contribute to the goals of FP7’s people programme, including Marie Curie Actions. ELIXIR’s training infrastructure will need to facilitate mobility of researchers – both within and between sectors and will also need to provide career opportunities for scientifically literate trainers. The success of ELIXIR’s training infrastructure will depend upon it working within the wider context of training for early-stage researchers, by helping them to develop the skills necessary for Europe to remain globally competitive.

5. Benefits of the ELIXIR training infrastructure

5.1. Contribution to European capacity building

The anticipated impact of ELIXIR’s training strategy is that it will make the benefits of ELIXIR accessible to Europe’s entire research community. This will permit the exploitation of Europe’s uniquely rich data collection by scientists both in academia and industry, facilitating translation of biological data into discoveries that will improve the quality of life of European citizens and increase Europe’s global competitiveness. ELIXIR’s training infrastructure will contribute technological development capacity in the European Research Area in the following ways:
• It will facilitate the full exploitation of public data in the life sciences by defining a training framework that catalyses knowledge transfer from the expert, through the trainer, to the end user.

• It will provide support for life science industries, to ensure that they can fully exploit available knowledge and so boost the knowledge economy.

• It will encourage the spread of excellence in computational biology throughout Europe, from the convergence regions to the outermost regions: e-learning resources and a network of centres of excellence in training will ensure that training is accessible to all.

Training needs to be incorporated into ELIXIR because:

• Optimal exploitation of life-science data is crucial to research, but without training the relevant users, this exploitation will never become a reality. Funding for training and outreach is often overlooked in favour of adding more functionality to bioinformatics resources, but the most feature-rich data resource is useless if people don’t know how to use it. A training infrastructure for ELIXIR will ensure that we bridge this gap.

• The flood of data arising from high-throughput technologies promises substantial and diverse increments in the well-being of Europe’s citizens, but many of the research scientists who are at the forefront of applied biology are at best unable to make full use of the data and at worst not even aware of Europe’s substantial collection of public data resources and tools. Outreach to these users through ELIXIR’s training infrastructure will fill this knowledge gap.

• Significant investment in ELIXIR will be wasted if the research community is not trained to use the data resources and tools.

• Effective training generates useful feedback that facilitates the future development of data resources.

• Training infrastructure costs are small compared with the cost of generating, collecting and organising the data, and yet are essential to translate data into knowledge.

• The already-huge user community is growing and diversifying relentlessly; ensuring that this community can make the most of the data is vital.

• New accession states are emerging as strong contributors to and users of the information; one of the most effective ways of making sure that they integrate effectively with the well-established member states is to raise awareness of the public data resources and support this new part of our user community to incorporate these resources into their research and into their science education.

5.1.1. Knowledge generation
Training is the means by which individual and community knowledge is developed. The ELIXIR training infrastructure will contribute to European knowledge generation by:

• Providing a framework by which trainers in the member states can incorporate bioinformatics user training into their own educational programmes.

• Stimulating interactions between bioinformaticians and experimental researchers, thereby increasing opportunities for the development of interdisciplinary approaches to research questions.

• Stimulating interactions between basic and applied scientists (for example, between bioinformaticians and researchers in pharmaceutical, agricultural or
environmental science) thereby accelerating the application of genomics and post-genomic approaches to health, a sustainable food supply, and conservation.

- Providing a ready means of collecting feedback from users which, in turn, will help to drive the future development of data resources.
- Adequate training underpins success in any profession; by providing a framework by which all life scientists can be trained to make the most of public data, we will contribute towards making Europe an attractive place for life-science researchers to live and work.

5.1.2. Industrial innovation

Industrial researchers are an important segment of ELIXIR’s user community. Our industrial users will play a vital role in ELIXIR’s vision, by being largely responsible for translating discoveries in basic biology into new technologies that will enhance the lives of European citizens and maintain Europe’s global competitiveness. Training for industrial bioinformaticians has been a core part of the EBI’s Industry Programme ever since its launch in 1997. Furthermore, industry has driven the development of many new data resources and standards in bioinformatics. Industry has already stimulated the development of new training formats: for example, the EBI’s e-learning portal was developed in response to demand from Small-to-Medium Enterprises.

ELIXIR’s training infrastructure will support industrial innovation in the following ways:

- The involvement of industry can help to shape the types of training modules offered, in terms of advising on what are the pressing needs for training. In this way, ELIXIR’s training strategy will help to bridge potential skills gaps, by ensuring that training is available in areas that are key for European competitiveness.
- There may also be an opportunity to work more closely with industry to develop bespoke in-house training modules; this would ensure that industrialists are getting the best out of ELIXIR’s core tools and resources, in a way that is relevant to their specific needs.
- ELIXIR will make bioinformatics user training accessible to all parts of the research community, including small companies with minimal resources for bioinformatics. This will give Europe the competitive edge by supporting innovation.
- ELIXIR will provide a precompetitive environment in which researchers from both academia and industry can come together with a shared goal and learn from each other.
- ELIXIR will provide ready access to training materials that, although not specifically aimed at particular educational programmes, can readily be adapted for this purpose. This will help educational establishments to respond rapidly to innovations in bioinformatics, thereby training a workforce with relevant skills for industrial research.

5.1.3. Societal impacts: contribution to the knowledge society

ELIXIR’s training infrastructure will contribute to the knowledge society by:

- Developing systems for sharing course materials, building on them, and then contributing improved materials, thereby building a self-improving knowledgebase for bioinformatics trainers.
- Developing systems for effectively distributing knowledge from a relatively small number of experts, through a larger number of trainers, to a very large number of end-users.
• Testing and, where appropriate, implementing new training methodologies and tools, thereby building pedagogic awareness and capacity within the life science community.

• Providing a means for ELIXIR’s training materials to be incorporated into educational programmes as educators see fit, rather than dictating how this should be done. This ‘bottom-up’ approach will help to drive appropriate innovation in science education at all levels, by people who know the needs of their students.

5.1.4. Independence and governance: securing European autonomy and knowledge

By building a sustainable infrastructure for biological data in Europe, ELIXIR will safeguard free access to biological data for Europe’s research community. But access alone is not enough: accessibility is key. ELIXIR’s training infrastructure is the vehicle by which access will be transformed into accessibility. The result will be a research community that not only has free access to global collections of life science data, but also knows how best to exploit this, and how to feed back to developers to ensure that Europe’s life science data resources continue to respond to the needs of research. In turn, this accessibility will help to drive excellence in European research.

5.2. Contributions to and synergies with national scientific activities/internal roadmaps in member states

ELIXIR’s training workpackage is already beginning to contribute to the efforts of national and pan-European training activities, through its initial efforts to collate and make readily available information on training activities throughout Europe. Our initial information-gathering exercise has identified some gaps that we are now poised to fill.

ELIXIR’s training infrastructure has great potential to connect the more interlinked bioinformatics resources of the established European member states with the relatively dispersed resources in the new member states. In Eastern Europe, for example, bioinformatics and related disciplines have traditionally been strong but have not yet harnessed the full potential of connections with centres of excellence in Western Europe. Thus far, the training of most bioinformatics users in Eastern Europe has been ad hoc and concerned largely with resources developed in the U.S. (e.g., at the NCBI). This has been, in part, due to a lack of structured training oriented towards the use of European bioinformatics resources. Thus, ELIXIR’s training strategy fits with the current efforts of some Eastern European countries (Poland, for example) to standardise the training in bioinformatics with that of Western Europe, to increase the use of Europe’s core data resources, and to integrate bioinformatics resources developed locally with a pan-European training programme, thereby increasing potential for spreading expertise throughout Europe.

The UK large facilities roadmap – a list of research facilities (national, European and Global) that may be eligible for capital funding in the UK - provides another illustration of how ELIXIR’s training strategy is closely aligned with priorities on a national level. The research councils are currently considering which large infrastructures in the biomedical and life sciences are of strategic importance to research in the UK. The need to access well organised, interlinked sources of biological data is shared by all of these infrastructures. The vast majority of emerging infrastructures listed in the roadmap specifically mention synergies with ELIXIR, and the existing ones, including facilities for mouse models, animal health, basic and translational medical research, are important data providers to, and end-users of,
Europe’s core data resources. ELIXIR therefore underpins the success of all of these infrastructures, and ELIXIR’s user base will comprise scientists who make use of these infrastructures. Other countries throughout Europe are developing their own roadmaps for infrastructure, with emphasis on areas of particular relevance for them. By improving accessibility to public-domain biological data, ELIXIR’s training strategy will enable these countries to realise their individual goals, including better understanding of individual populations and the use of this knowledge to improve healthcare, and better understanding of crops that are of great economic importance to particular countries, leading to improved crop production.

5.3. Contributions to and synergies with other ESFRI projects

ELIXIR will support at least nine ESFRI biomedical and environmental science infrastructures. Each of these has its own training strategy. ELIXIR’s training infrastructure will be directly relevant to each of these, and we envisage that course materials developed by the ELIXIR training infrastructure will be incorporated into the training programmes developed by these BMS projects, each of which represents a distinct user group for ELIXIR. A pan-European framework for training these different types of user will be of strategic importance to each of them, helping them to realise benefits such as:

- Integrating different types of molecular data to model and understand biological systems
- Linking from biobank samples to relevant genomic information
- Mining the core databases and the scientific literature for potential drug targets
- Annotating the human genome with functional information from large-scale mouse genetics experiments
- Linking from genotype to phenotype and using this information to understand disease susceptibility and/or drug sensitivity
- Understanding biodiversity by linking taxonomic, environmental and genomic data
- Understanding biological function by combining macromolecular structural data with information on sequence, domain structure and binding sites
- Linking information on biological macromolecules structure–activity relationships of drugs and drug-like molecules to develop new drugs.
- Mining metagenomics data sets for information on biodiversity.

A relevant case in point is the Innovative Medicines Initiative (IMI) Education and training pillar. IMI is a public-private Partnership between the pharmaceutical industry (represented by the European Federation of Pharmaceutical Industries and Associations; EFPIA) and the European Communities represented by the European Commission. IMI’s overall goal is to make Europe again the world leader in pharmaceutical research for the benefit of the economy and society, by removing research bottlenecks in the current drug development process. Education and training is one of IMI’s four focus areas, and the six most advanced ESFRI BMS infrastructures are now in the final stages of contract negotiation to jointly coordinate IMI Call Topic 14, to establish a network (EMTRAIN, www.emtrain.eu) to facilitate and coordinate European training and education relevant for stakeholders of medicines research and development. ELIXIR, together with representatives from EFPIA, will co-lead two of EMTRAIN’s workpackages, one on supporting continuing professional development throughout medicines development, and another on new
learning methodologies. Both of these workpackages are of direct relevance to ELIXIR’s own training goals; EMTRAIN will benefit from ELIXIR’s mapping and gap analysis, whilst ELIXIR will readily be able to make contact with emerging user groups, thereby helping to shape the future of bioinformatics training for medicines development.

5.4. Potential benefits to stakeholder communities

Training is the most effective way to ensure that the potential benefits of ELIXIR are actually realised. ELIXIR is concerned with providing a service to users of Europe’s core data resources, but providing this service without complementary training will guarantee the failure of ELIXIR by failing to empower its users to exploit the data.

5.4.1. Benefits to end users, power users, trainers and educators

The ELIXIR training infrastructure allows for expert knowledge, generated and updated by training support staff embedded within bioinformatics service teams, to rapidly be transferred to trainers, modified if necessary for their own applications, and then transferred to users.

By supporting the production of training materials that are modular and agnostic of course format, ELIXIR will be able to supply the materials that trainers and educators need, without dictating how they teach. Trainers develop their own courses and course collections for good reason and we have no interest in stifling this creativity, which is essential to effective training.

5.4.2. Benefits to data resource and tool providers

ELIXIR’s training infrastructure will provide a ready means for feedback to the data resource and tool developers, thereby facilitating the further development of data resources that are tailored to users’ needs.

The funding mechanism that we recommend will embed training support into Europe’s major data resources and tools. By having dedicated training support staff within their teams, database and tool providers will be able to focus on their core business of developing and maintaining their resources, whilst retaining that all-important link with their users.

5.4.3. Benefits to funders

ELIXIR’s training infrastructure will help funders to realise the full benefit of their investment, by ensuring that Europe’s life-science research community is fully empowered to use the publicly available data resources and tools.

6. Steps for realising the benefits

Steps for realising the benefits of ELIXIR are detailed in section 4.5. After our initial survey of the state of the art (detailed in section 4.1) our first step will be to build a community of training stakeholders that will ensure that ELIXIR’s training infrastructure meets their needs. This community will help us to complete our mapping and gap analysis, and to begin to build effective resources for sharing training information, and course content, throughout Europe.

A centralised organisational body, which takes responsibility for engaging with these stakeholder groups, will be vital to the success of ELIXIR’s training initiatives. This will help to mobilise existing training networks in the member states and promote the sharing of best working practice. Equally important is funding embedded within each of ELIXIR’s data resources and tools. This will ensure that the resources themselves take responsibility for providing effective and up to date training information.
With this two-pronged approach, it will be possible to vastly improve accessibility to Europe’s life science data. This, in turn, will enable Europe’s research community to capitalise fully on the investments that funders have already made in gathering the data, and will enable Europe to transform scientific understanding into better quality of life for Europe’s citizens.
Appendix 1: Examples of Current bioinformatics training programmes

The training programmes summarised below are ones that members of the ELIXIR training committee have been closely involved in setting up and implementing. In the course of discussing and documenting these examples, it became very clear to us that each programme has its own merits, and that there is no single ‘right’ way to approach bioinformatics user training. The analysis of these programmes led us to conclude that it would be unwise for ELIXIR to enforce a single, standardised user-training methodology. Instead, we have focused on how ELIXIR might solve problems that we all encounter on a regular basis, such as how to keep training materials up to date with rapidly evolving data resources and how to ensure that trainees enroll on courses that are best suited to them. We detail our different approaches to user training below.

Face-to-face training

The BioSapiens European School of Bioinformatics

The BioSapiens European School of Bioinformatics was held twice a year for five years (2004-2009). Specific aspects of the school are:

• It was held each time in a different country and directed mostly to bench biologists or newcomers to bioinformatics.

• It is essentially practical. Each day there includes no more than 1-2 hours of lectures, the rest is practical work.

• Each day was dedicated to one topic. Topics might differ slightly in different editions, but the general schema was Databases / Sequence analysis / Structure analysis / Microarray data analysis / Systems biology. The last topic was introduced following a very successful collaboration with the EU-funded ENFIN Network of Excellence.

• The instructors were post-doctoral researchers participating in the research aspects of the BioSapiens network (and of ENFIN for the last day).

• Each instructor was in charge of one day and supported at least one other instructor on a different day.

• Each school had 40-50 participants. There was typically at least one computer between two participants. Experience has shown that working in pairs is much more effective than working alone.

• At the end of the school, participants were asked to complete a questionnaire.

Our assessment of the results of this exercise is extremely positive, and this is well testified by the constant high number of applicants and by the questionnaire results (available on the BioSapiens web site). One added value of the school was that it fostered close interaction between the post-docs acting as instructors. For young researchers, any opportunity to build European network of collaborators is of enormous value for their future career.

We believe that one ingredient that made the school so successful was the financial support (through BioSapiens), which allowed a large participation of students. Another was the fact that instructors were young active researchers, both ready to dedicate a substantial amount of their time to the preparation of the practicals and aware of the last developments in the resource.
It should be mentioned, as a proof of the relevance of training for both end users and developers, that the input of the school participants greatly helped improving and benchmarking the DAS portal of the Biosapiens NoE and thereby contributed to the development of an important pan-European resource.

**The EMBL-EBI hands-on user training programme**

http://www.ebi.ac.uk/training/handson/

Trainees come to the EBI to learn how to use Europe’s core data resources and some of the tools associated with them. Courses are typically 2–5 days in length and take place in a purpose-built IT training room, in which each trainee has his/her own desktop computer. The curators and/or software developers who produce the data resources devise the programme for each course, with significant input and support from a dedicated training team. If appropriate (and if funding allows), we also invite relevant external lecturers.

The formal programme comprises a mixture of lectures and hands-on practice sessions. All lecture materials and hands-on tutorials are provided in a course handbook, and are made available electronically to all delegates who complete an evaluation form after the course.

The trainers are strongly encouraged to integrate with the trainees at mealtimes and coffee breaks. When courses are fully residential, the trainers and course organisers also have evening meals with the trainees. This provides plenty of opportunity for the trainees to discuss their own projects with the trainers.

The trainees are asked to complete an online evaluation form at the end of each course (see section on evaluation). The evaluations are analysed and summarised by a member of the training team, who circulates this analysis to all trainers. Whenever possible, we hold debriefing sessions shortly after each course in which we discuss the strengths and weaknesses of the programme and any technical hitches. These discussions are used to make a decision as to whether the course should be held in the future, and to action any improvements to the programme in general.

**The Bioinformatics Roadshow**

http://www.ebi.ac.uk/training/roadshow/

Originally organised under the auspices of the EU-funded FELICS Integrated Infrastructure Initiative and now funded by the SLING Integrating Action, the Bioinformatics Roadshow is coordinated by the European Bioinformatics Institute in collaboration with our FELICS/SLING partners: the Swiss Institute of Bioinformatics (SIB), the European Patent Office and the BRENDA project. The roadshow provides a travelling training programme covering Europe’s most widely used biological data resources, which include the EPO’s patent databases, SIB’s data resources and BRENDA as well as the EBI’s core data resources.

The Bioinformatics Roadshow aims to address the needs of European scientists for hands-on bioinformatics training. Like the hands-on programme, it combines presentations and hands-on practical sessions in which experts in key bioinformatics applications guide users through their selected topics.

Roadshows are typically held over two days (this is decided by the host) and are modular in nature. The roadshow ‘modules’ conform to ELIXIR’s concept of a training module, in that they can be mixed and matched to create bespoke courses suitable for the host institute. If space allows, sessions can be run in parallel to cater to the differing interests and requirements of the audience.

The programme for each roadshow is devised by the host in collaboration with the EBI’s training team, and may also involve individual discussions between the host
and the trainers to define the host institute’s training needs in more detail. A date is
fixed to coincide with trainer availability (this is increasingly challenging as demand
for trainers is high and the vast majority of our trainers are full-time curators and
software developers who make time to train). The host is expected to promote the
roadshow within his/her organisation (and beyond it if appropriate), and the EBI
provides materials and support to help the host do this. Under FELICS, the host was
also expected to pay the travel, subsistence and accommodation costs of the
trainers, to provide a suitable IT training room, and to print out the training materials,
which we provide electronically to obviate shipping costs. Under SLING, we are very
fortunate to have funding for trainer travel and subsistence; this increases the reach
of the roadshow to institutes that could not have afforded to host a roadshow under
FELICS.

The Gulbenkian Training Programme in Bioinformatics
http://gtpb.igc.gulbenkian.pt/bicourses/

The Gulbenkian Training Programme in Bioinformatics GTPB) was created in 1999 to
respond to the need to provide user training on a more regular basis. Prior to this, the
Instituto Gulbenkian de Ciência had been organising courses on demand since 1986.
The GTPB organizes 10-15 international training courses in Bioinformatics per year.
These is a core set of themes that is repeated once every year and an additional set
of thematic courses that are introduced anew, mainly in response to user demand.
Every year there are always one or two introductory courses (sequence analysis).

Teachers (trainers) are recruited from a pool of people that accept our challenge and
respond positively to a set of requirements, the main on being the ability to present a
good plan, months in advance. The plan is discussed with the Organiser (Pedro
Fernandes) and tailored to known requirements that are derived mainly from user
feedback.

Attendance is selected from formal applications received 30-35 days before each
course. Applicants need to present a mini-CV and a one-paragraph letter of intent (I
wish to attend this course because…)

All courses are taught and documented in English. Our attendance is rather
international.

The courses are taught in an intensive manner, 7-8h per day in a dedicated training
room. There are 10 tables, each with a computer and 2 seats. The 11th computer is
for the teacher and projects on a large size screen. A dedicated server in the room
provides several services including access to a shared disk space that is conveniently
used for communication with the audience, for the quick and easy deployment of
materials and for sharing results in the classroom.

Experience has shown that key issues for fast acquisition of skills include:

• That attendees work in pairs.
• That interruption for clarifications in any class or practical is possible at all times.
• That attendees take only minimal notes, as they know that each participant takes
  home a CDROM with all the materials at the end of each course.

Coffee breaks are 30 minutes long, allowing for useful discussions to occur.

Costs: The average duration of a GTPB course is 4 days and the average cost is 500
Euro. The course fee is Euro 80 per day. Subsistence in the area is quite easy. The
absolute minimum of Eur 28 per day has been achieved in 2008. We have access to
accommodation in a local academic residence for Euro 21 per night in a single room
with private bathroom. Food can be very cheap too. So, for a 5-day course, each
attendee needs to spend a minimum of Euro 550 locally, plus travel costs.
e-learning

Teacher-led e-learning at the Center for Biological Sequence Analysis, Technical University of Denmark
http://www.cbs.dtu.dk/courses.php

At the Center for Biological Sequence Analysis (CBS) a method has been developed that combines classical face-to-face training with e-learning. The method aims at offering courses to the largest possible audience while at the same time keeping the added cost and effort of e-learning to a minimum. Here, training sessions are simply broadcast versions of the classroom teaching with the addition of online students. In the standard setup, the broadcast contains audio/video of the teacher, the slide show (if any) and a chat room. The online participants have access to the teacher via the chat interface where a moderator will answer questions or pose them to the teacher during lessons. The moderator thus effectively acts as the voice of the online students in the classroom. Adobe Acrobat Connect Professional has been used at CBS, but most newer videoconferencing tools can be used. While simultaneous teaching of online and onsite students does benefit from a dedicated room with cameras and sound system, less advanced setups will do.

In terms of workload and personnel, the added cost of this approach compared to classical face-to-face training consists of setting up the lessons beforehand and having a chat moderator present during lessons. From the trainer’s point of view, the added trouble of broadcasting the sessions is thus minimal and this solution is therefore attractive in cases where external teachers are used, since they do not need to be familiar with the system. From the student’s point of view, using the system is equally simple – the only requirement is access to the Internet via any standard web browser. The flexibility of the system goes beyond simple broadcasting of classroom teaching in that it is possible to teach from a remote location. In this case, lectures are broadcast to online students as well as the classroom students. Again, the only requirement for the teacher is access to a standard web browser and a webcam/camera and microphone. This makes inviting guest lecturers possible in cases where travel time or expenses would otherwise prohibit it.

Teacher-led e-learning at CSC - IT Center for Science Ltd.
http://www.csc.fi

CSC’s training facility consists of a classroom and multipurpose studio to arrange courses and events together with CSC. CSC has a specifically equipped classroom with twelve PC workstations and a PC for the teacher. The computers have Windows XP and Linux user environments and a network connection. In addition, the equipment includes normal audio-visual devices, a digital projector and a document camera. The multipurpose studio at CSC is equipped with state-of-the-art multimedia technology. It is a 7m x 7m room with studio quality acoustics and mostly black interior. The projectors on a soundproof space can project wall-sized video or computer presentations, such as Access Grid connections.

A bioinformatics course at CSC typically consists of lectures and hands-on exercises with the emphasis heavily on exercises. We also try to allow students to use their own data on the course, if they have such. Videos can be recorded on the fly alongside with the course. All course material (slides, exercises and videos, http://www.csc.fi/english/csc/courses/archive) is usually made publicly available from the CSC site. Materials are not classified according to specific topics, but bioinformatics course videos in the archive are grouped under sciences and methods sections on the CSC web pages. Sometimes the videos are used afterwards on other courses as part of the curriculum. Several institutes abroad are also known to use
our videos in this way. The tendency in bioinformatics education is towards shorter
videos and better pedagogy with, e.g., videos alternating with student activities.

Overall, in the year 2008 CSC arranged 81 courses and events (145 course days).
Altogether, over 2000 people were registered to these. The number of participants in
bioinformatics courses was 164 on site and over 300 including off-site participants.
Bioinformatics was one of the most active training areas, focusing on microarray data
and sequence analysis and protein modelling. Other focuses included chemistry,
computing technologies and data administration.

EMBER
http://www.ember.man.ac.uk

Manchester University led a European consortium (with 10 partners) to develop a
European Multimedia Bioinformatics Educational Resource (EMBER): the result was
a Web-based introductory bioinformatics course and the same stand-alone course on
CD-ROM. The former is freely accessible online; the latter was made available for

A key concept embodied within EMBER is that the course is interactive - in other
words, that students should learn both by doing and by reflecting on what they've
done, rather than simply by reading about the theory via dry Web pages. The course
contains a number of 'chapters', the first six of which constitute the 'Basic tutorial';
the next three progress to more advanced topics; and the final chapters offer a
number of 'Case studies'. Supporting the chapters are introductory aims, which state
the learning objectives; step-by-step instructions on what to do; structured,
background theory, with diagrams, images and so on; end-of-chapter quizzes, which
help students to understand what they got right and what they got wrong; a
comprehensive glossary; a list of references; and a feedback form.

Behind EMBER is a database management system that allows course coordinators
to set up courses with their own unique identifiers; this means that the progress of
particular student cohorts can be monitored during the course, and their success in
the quizzes before and at the end of the course assessed. Thus, there is in-built
evaluation, both for students and for course coordinators.

Although designed to be a self-contained interactive tutorial in bioinformatics,
EMBER is now used in a variety of different contexts: (i) as the practical component
of traditional face-to-face training courses (including bioinformatics summer schools
in the UK and abroad, as part of roadshows, continuing professional development
programmes, etc.) and Masters modules; (ii) as a self-contained module on distance
learning Masters programmes (e.g., at Manchester and Birkbeck College) and
Graduate Training Programmes; and (iii) as a self-contained, standalone module on
the Web, freely accessible by the community, for individual self-paced study or
inclusion within courses worldwide. EMBER is currently also being piloted as an in-
house training vehicle at AstraZeneca. At the University, the entire system is also
being imported into Blackboard, in order to allow us to be able to customise its
content more readily.

For more details, there is an article describing EMBER in the online journal, Bee-J:
www.bioscience.heacademy.ac.uk/journal/vol6/Beej-6-4.htm

The EBI's e-learning pilot project: trainee-led e-learning
www.ebi.ac.uk/training/e-learningcentral

The EBI has developed an e-learning portal to complement its face-to-face training
programmes. We have taken a different approach to DTU and the CSC, offering self-
paced courses that can be completed in the absence of a trainer. The content of the
courses closely reflects that of the modules that we teach as part of our hands-on
courses and roadshows. Each course contains a number of modules, and each
module contains some or all of the following:

- Step-by-step print tutorials (PDF format).
- Video tutorials (most suitable for conceptual materials that don’t need updating
  on a regular basis).
- Self-marking ‘key concepts’ quizzes that provide explanations when the trainee
  selects an incorrect answer.
- Reflective tasks: problems for the trainee to solve using the concepts that they
  have just learned. Hints are provided if the trainee is unsure how to get started on
  each task.

The e-learning portal is built using Moodle, an open-source e-learning platform that is
used by many schools and universities in Europe.

Like the University of Manchester, we have found that our e-learning portal is well
suited to providing an adjunct to face-to-face training. It is also a useful platform for
trainers to exchange and compare course content, methodologies and ideas. As
technology improves, we anticipate that we’ll be able to build more interactive and
time-sensitive content into the portal; this will, we hope, obviate the updating
bottleneck and improve the dynamism of the site, encouraging trainees to attend
one-off events.

eProxemis
http://www.eproxemis.com/

The Swiss Institute of Bioinformatics developed an e-learning portal – EPROXEMIS –
which is based on a purpose-built e-learning platform. Owing to lack of funds for the
continued development of their platform, they have investigated the possibility of
porting their content to Moodle. Their conclusion is that although Moodle is very
popular and user friendly, it is not ideal for self-learning content because it is more
oriented towards teacher-led e-learning, in which interaction with an instructor forms
an integral part of each course. This cautions against the ELIXIR training committee
making recommendations as to the type of platform ELIXIR uses for e-learning.
Appendix 2: thoughts on benchmarking

Early on during our discussions we agreed that although there is much demand for a repository of training materials, such a repository would be of little use unless its content is carefully categorised and benchmarked. This appendix summarises our thoughts on how ELIXIR might approach the benchmarking challenge.

The distribution platform for ELIXIR training materials should be compared with existing solutions and organizations focusing on sharing digital learning resources. Benchmarking should include a survey of existing technical solutions to find out whether they are appropriate for, or could be modified to make them appropriate for, a pan-European bioinformatics user-training platform. Two exemplary systems for distribution of training materials in the spirit of Creative Commons philosophy include the Connexions system for open learning materials (http://cnx.org/) and Merlot, a collection of peer-reviewed higher education online learning materials (http://merlot.org/). Any system that ELIXIR builds for bioinformatics could be benchmarked against these or similar systems.

The best practical solution in Finland known to us currently are processes defined by the ApuMatti (http://apumatti.helsinki.fi/) wizard for publishing digital learning material or topic-case driven methodology (http://urn.fi/URN:ISBN:951-39-2264-2).

There is also international cooperation in the field of open e-learning materials. The Open Archives Initiative (http://www.openarchives.org/) develops and promotes interoperability standards that aim to facilitate the efficient dissemination of content (e.g. Protocol for Metadata Harvesting OAI-PMH). Further, EdReNe (Educational Repositories Network) has also published a report on the current status of educational repositories in Europe (http://edrene.org/results/currentState/index.html).

ELIXIR could also consider a benchmarking exercise to identify the most effective e-learning methodologies for bioinformatics user training, by comparing training models and usage statistics among participating organisations.
Appendix 3: thoughts on monitoring success of training courses

Perhaps the best judges of whether a course has achieved its aims are the trainees themselves. When coordinating training efforts more broadly, it is also essential to seek feedback from the trainers and, if applicable, local hosts. However, gathering feedback, by itself, is only half of the story. Systems must be put in place to incorporate feedback into future courses. One valuable function that the ELIXIR Training Support Unit could perform would be to collate best working practice on evaluations and devise a set of core questions for incorporation into all bioinformatics user-training evaluation forms. This would still allow trainers to ask their own specific questions.

The FELICS I3 developed a three-pronged approach to evaluating its roadshow programme that is now being used under the SLING Integrating Action:

Feedback from trainees: towards the end of each roadshow, time is incorporated into the programme for the trainees to complete an online feedback form [see, for example, the form at http://www.surveymonkey.com/s.aspx?sm=hMS4L5kxU58_2bvMHjEIQ75A_3d_3d]. They are incentivised to do this by being provided with a password to gain access to the training materials online only when they have completed the evaluation form.

Feedback from roadshow hosts: the roadshow hosts are also asked to complete a very simple one-page form. This allows us to determine whether the roadshow met the goals of the host organisation as well as those of the individual trainees. This is increasingly being done via phone interview between the host and an EBI training coordinator.

The trainers are asked to complete a one-page form that helps us to determine the suitability of the venue for future training courses, and also helps to identify any organisational glitches so that we can prevent them from happening in the future.

As soon as possible after each course, these three types of evaluation are reviewed by the roadshow’s coordinators. A summary of the trainees’ feedback is sent to the trainers so that they can incorporate any suggestions for improvement into future courses. Any feedback of a more general nature is addressed by the roadshow coordinators. If there are any organisational issues, these are discussed, and workable solutions are found. It has proved very challenging to meet with all the trainers after each roadshow, but if any significant issues arise we discuss solutions to these by e-mail.

This approach to feedback enabled us to make significant improvements to the roadshow programme throughout its first two years of operation. These included:

• providing a robust way of ensuring that training materials are issued to all trainees
• providing guidelines to our hosts to support their efforts to promote their roadshow locally
• Developing a centrally run registration system so that we can guarantee significant attendance and provide the hosts with good value for money
• Ensuring that our hosts provide adequate facilities for our trainers.

CSC uses a standard set of questions to collect feedback on all courses (https://www.webropol.com/P.aspx?id=299584&cid=80519478). The questions are designed to be generic enough to fit almost any course or event.
Perhaps the best method for measuring training effectiveness is to test how much the trainees have learned; however, this has its setbacks in a face-to-face training environment as it can affect the convivial and open atmosphere that we try to foster; furthermore, we do not have a baseline as it would be unusual to test the trainees’ knowledge before each course. The EBI has some limited experience of evaluating the effectiveness of training courses by testing its trainees. We have developed a simple 10-question quiz that trainees attending the sequence searching module of the 8th BioSapiens European School in Bioinformatics were asked to complete. 57% of the delegates achieved a grade of 83% or above. We are using the same test to evaluate users of our sequence searching e-learning course, as part of the ELIXIR pilot study on training methodologies.

Within the EMBER virtual learning environment, we provide an ‘initial quiz’ comprising 24 questions that cover all of the chapters offered within the resource, each chapter then has its own ‘end-of-chapter quiz’, and, before leaving the system, a ‘final quiz’ is also provided. There are several benefits to this kind of system: (i) the initial quiz gives users instant feedback both on areas in which they’re confident and on areas that need improvement, before they start the tutorial; (ii) the end-of-chapter quizzes give users step-wise feedback as they progress through the tutorial (the quizzes are designed such that they inform students not just that particular answers were right or wrong, but why they were right or wrong - and if students get lots of wrong answers, the quizzes encourage them to go back and revise or re-do the relevant parts of the chapter); (iii) the final quiz gives users feedback on their overall score at the end of the tutorial, so they can see how much they have improved; and iv) the collation of quiz results in EMBER’s database gives course coordinators both an overview of student performance and relative improvement, and an illustration of which parts of the tutorial and/or which questions within the quizzes are causing problems (e.g., in terms of which are too easy, which are too difficult). Potentially, then, this is a self-improving process, whereby students can use the results of the quizzes to evaluate their progression, and coordinators can use the results to evaluate the effectiveness both of the material and of the quiz questions themselves. A feedback form is also used to capture student reactions to the experience of using the VLE itself, and how useful they found the tutorial overall.