

ELIXIR-UK

the UK Node of the

European Life Sciences Infrastructure for Biological Information

www.elixir-europe.org

elixir-uk.org



The USA

McKinsey & Company

McKinsey Global Institute



June 2011

Big data: The next frontier
for innovation, competition,
and productivity



Big data—capturing its value

\$300 billion

potential annual value to US health care—more than double the total annual health care spending in Spain

€250 billion

potential annual value to Europe's public sector administration—more than GDP of Greece

\$600 billion

potential annual consumer surplus from using personal location data globally

60% potential increase in retailers' operating margins possible with big data

140,000–190,000

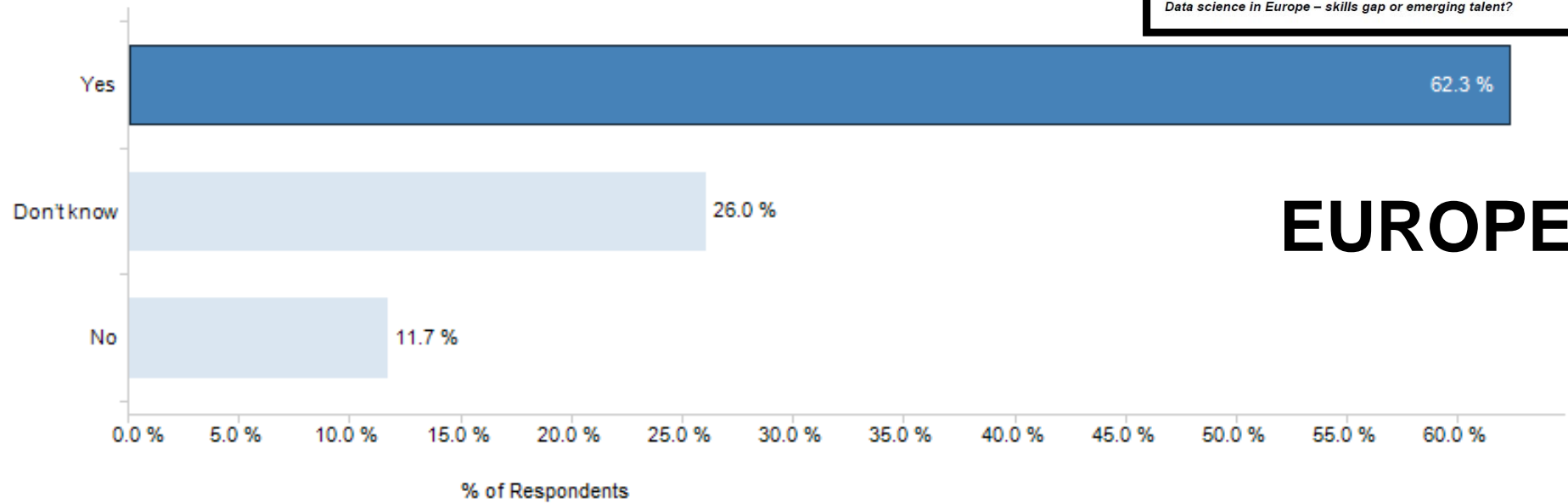
more deep analytical talent positions, and

1.5 million

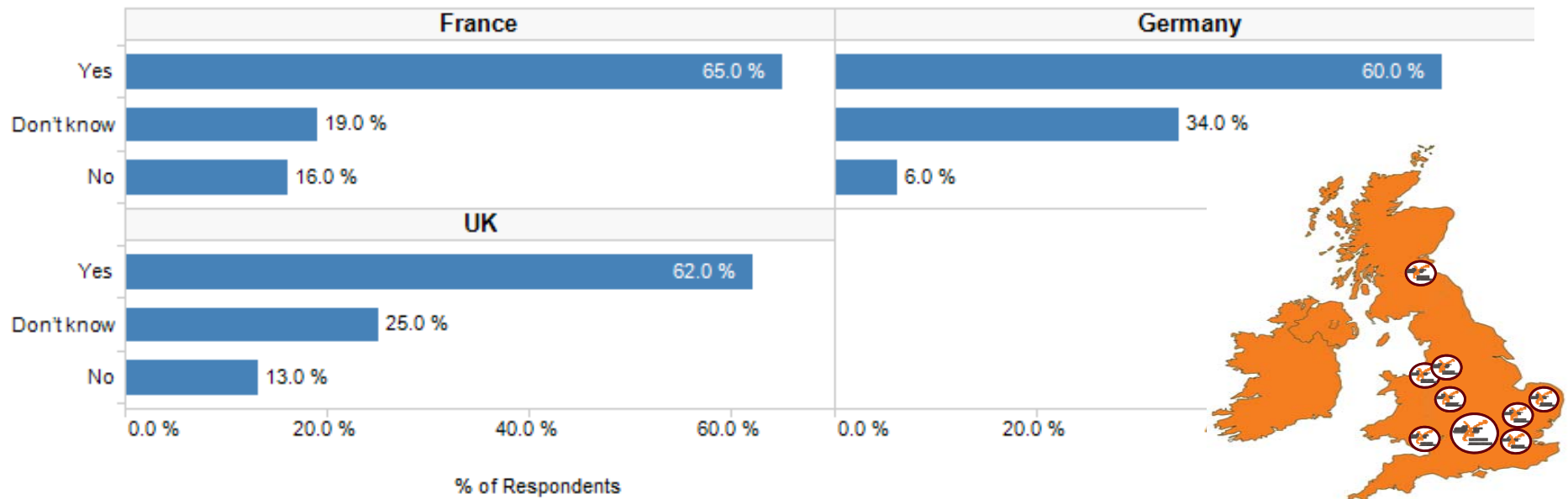
more data-savvy managers needed to take full advantage of big data in the United States

Q3. Do you believe there is a skills shortage in the area of data science/big data/data analytics?

Teradata research project
Data science in Europe – skills gap or emerging talent?



EUROPE



UK: Poor data quality hindering government open data programme

- “A *Computer Weekly* analysis of 50 spending data releases by the Cabinet Office since May 2010 has shown they were so marred by “dirty data” and inconsistent computer encoding, systematic scrutiny would require advanced computer programming skills.”

Thursday 28 August 2014

<http://www.computerweekly.com/news/2240227682/Poor-data-quality-hindering-government-open-data-transparency-programme>

Establishing a fresh UK activity

Oxford University Computational Genomics Analysis and Training (CGAT)

The University of Manchester

The Oxford e-Research Centre

European Bioinformatics Institute (EMBL-EBI)

University of Cardiff & NERC EOS Centre

The Genome Analysis Centre (TGAC)

University College London

University of Birmingham

University of Edinburgh

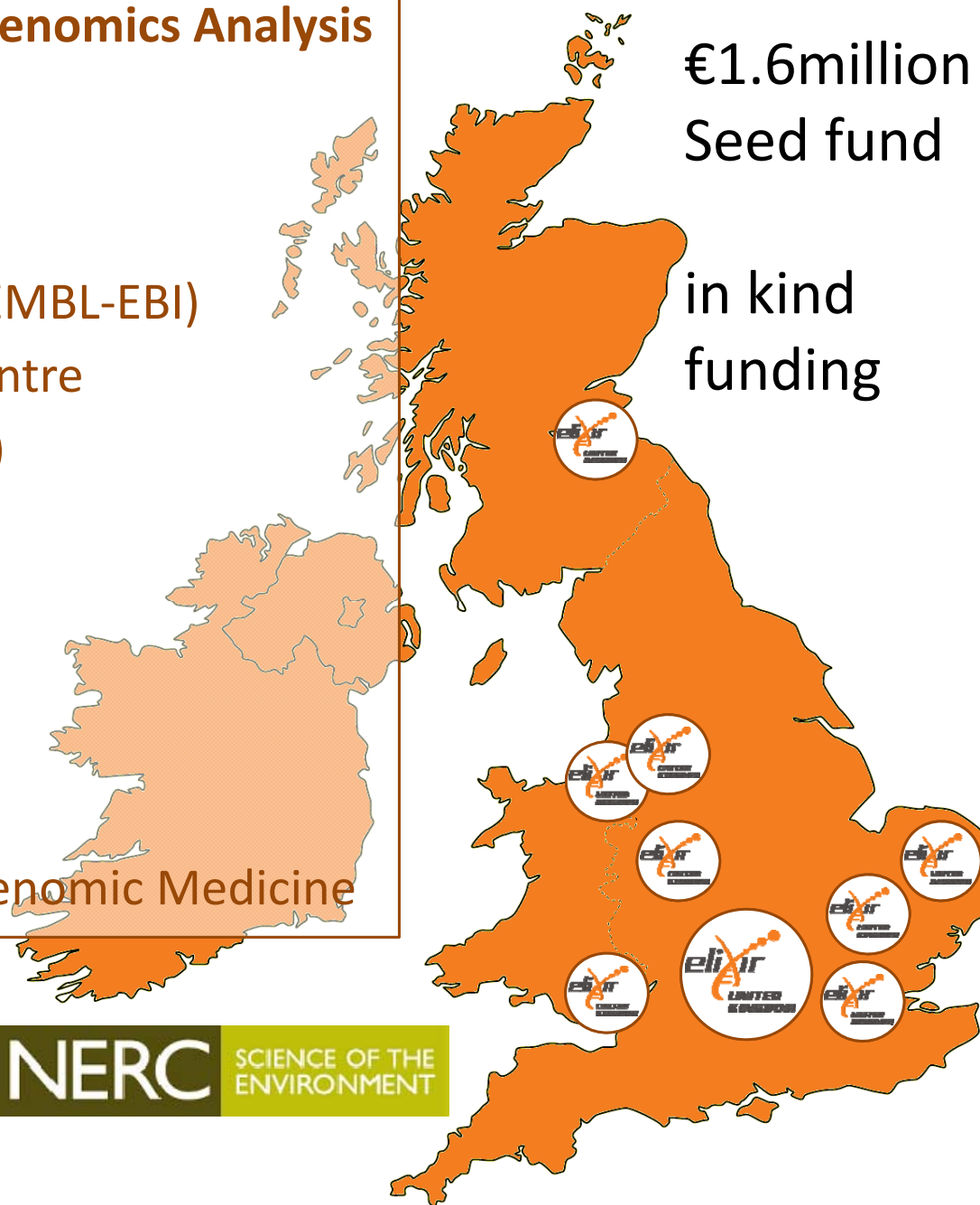
Queen Mary, London

University of Cambridge

University of Liverpool Centre for Genomic Medicine

€1.6million
Seed fund

in kind
funding

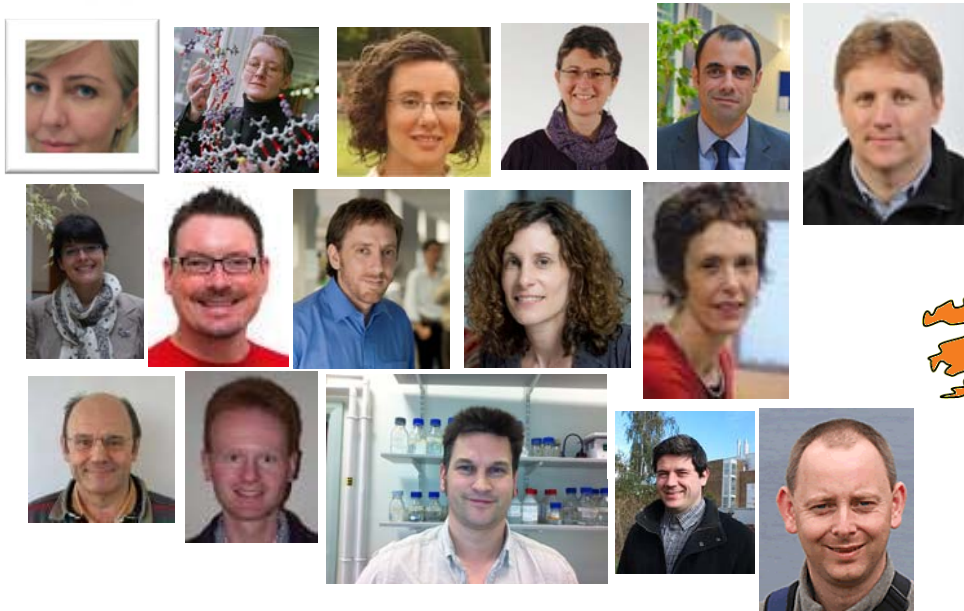


Harness existing expertise

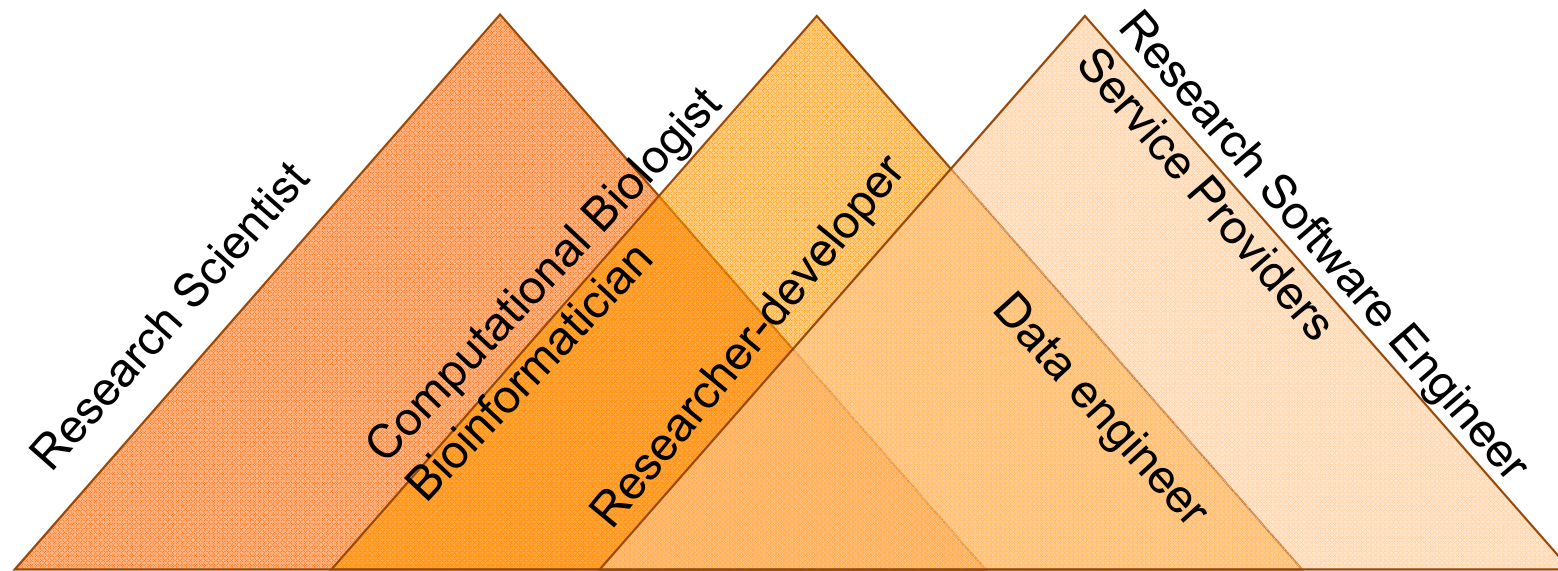
Core Staff



TeSS Staff



Train across the Spectrum



Life Science
Researchers

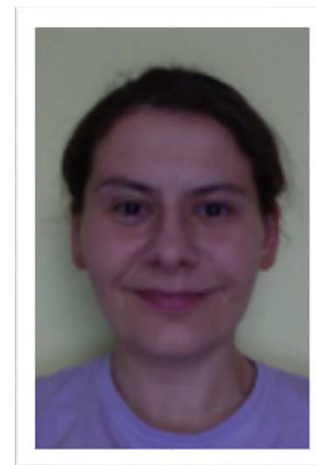


Technical
Infrastructure
Service Providers

TCRS
Training
Coordinator,
Research
Science



Lee Larcombe



TCIT
Training
Coordinator,
Infrastructure
Technology

Aleks Pawlik

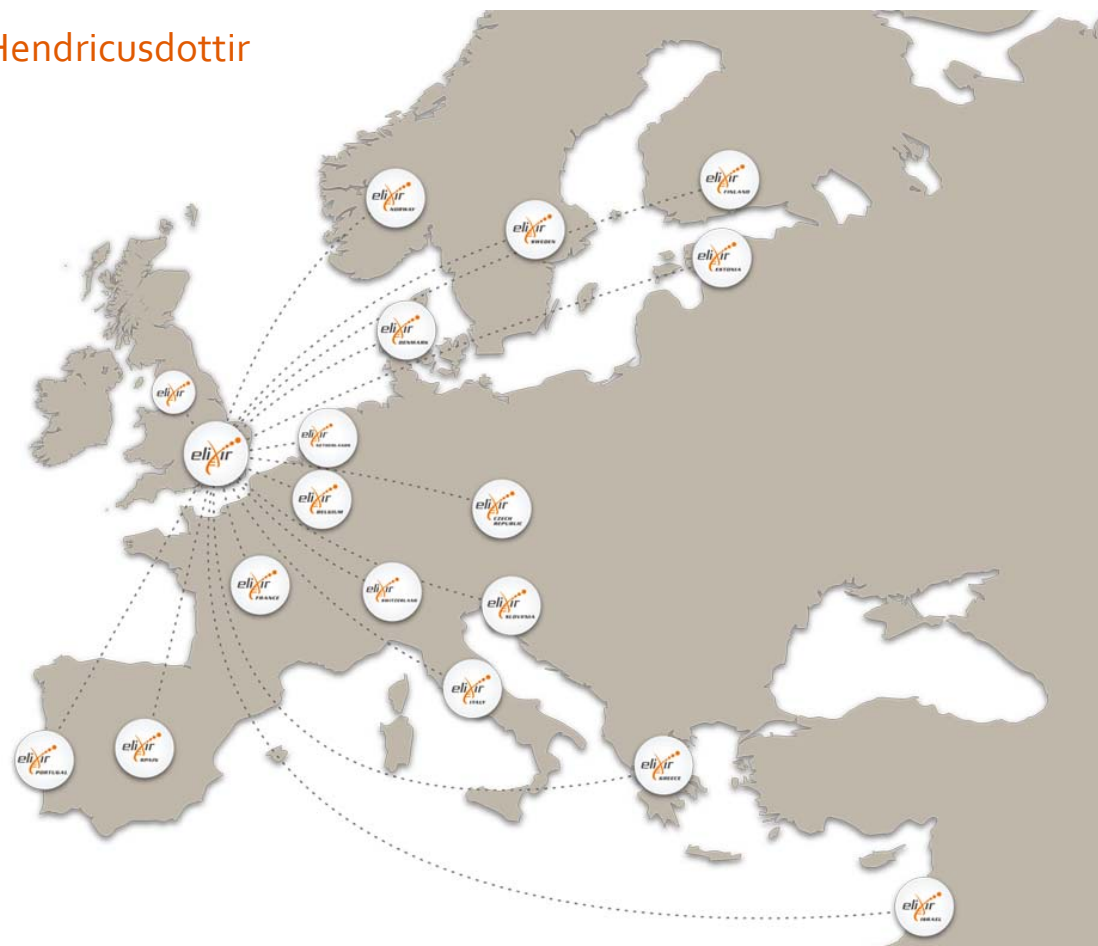
ELIXIR-UK working across Europe: Training Coordination Group

Mission: to establish an interacting ELIXIR wide training community & to ensure coherency in the delivery of training related to ELIXIR activities.

TrCG members:

BE	Katrijn Vannerum
CZ	Daniel Svozil
DK	Peter Longreen
EE	Hedi Peterson
FI	Eija Korpelainen
FR	Julie Thompson
IL	Michal Linial
IT	Allegra Via
NL	Celia van Gelder
NO	Ståle Nygård
PT	Pedro Fernandes
SI	Brane L. Leskosek & Peter Juvan
ES	Oswaldo Trelles
SE	Sara Light
CH	Patricia Palagi
UK	Rita Hendricusdottir & Lee Larcombe
EMBL-EBI	Sarah Morgan

Chair: Rita Hendricusdottir



TrCG facilitating training in UK/Europe

Form a strong training community within ELIXIR



Partnership with training professionals

Enhance quality of training

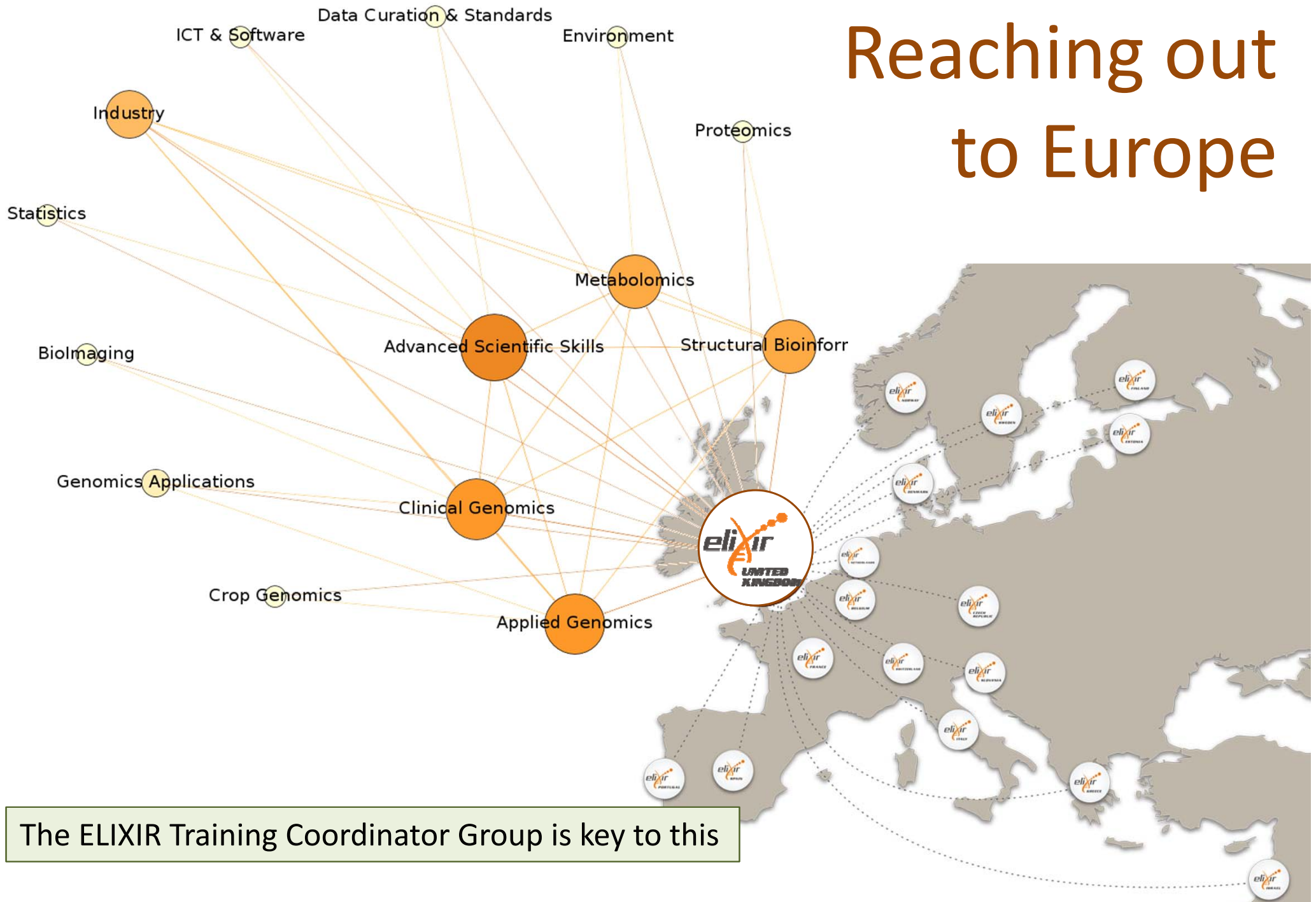
Enhance impact of training

Funding opportunities

Promote UK training in Europe

- Form strong training community within ELIXIR
- Drive partnerships
- Coordinate development of training
 - Training/workshops: Related to ELIXIR infrastructure
 - Train the trainers: Universities and Industry
 - E-learning: improve E-learning training
- Increase accessibility for training
 - TeSS: ELIXIR training portal
- **11 Nov 2014: First TrCG Face to Face meeting**
- **EXCELERATE Training WP.**

Reaching out to Europe



The ELIXIR Training Coordinator Group is key to this

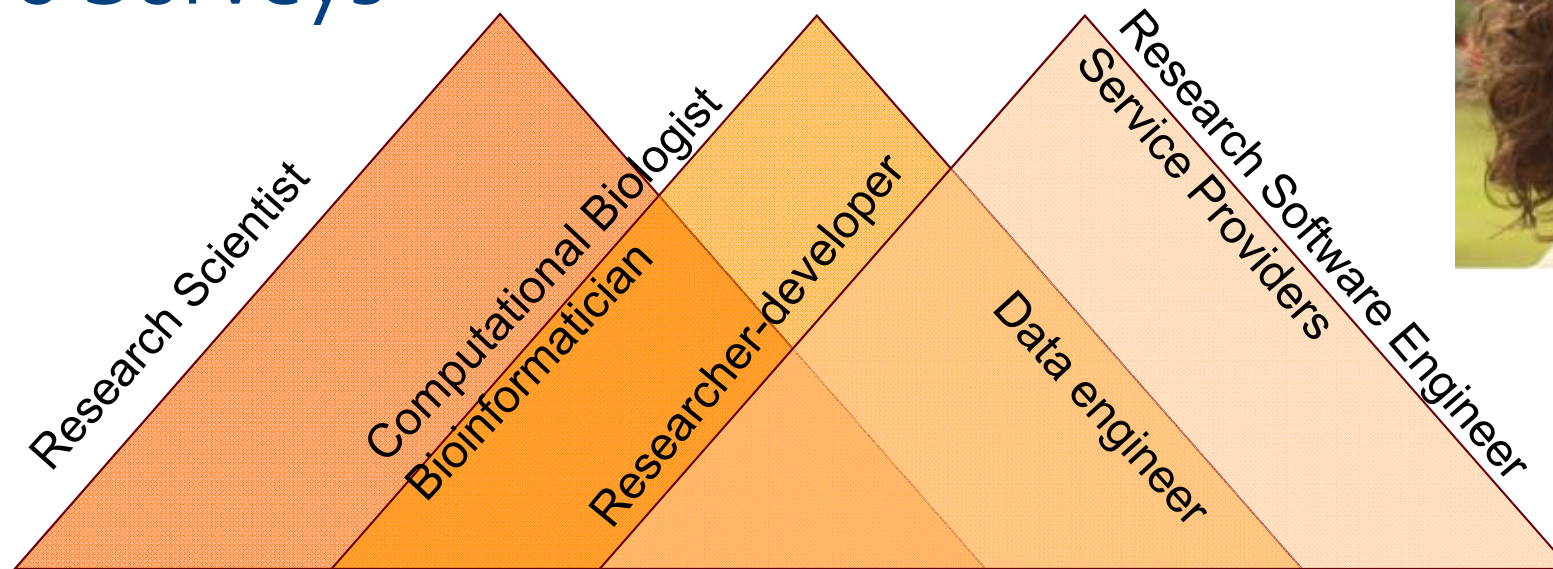
ENGAGING WITH INDUSTRY

Appointed an Industry Engagement advisory committee

Confirmed members:

- Claus Bendtsen, AstraZeneca
- Mark Forster, Syngenta
- Samiul Hasan, GSK
- Wendy Filsell, Uniliver
- William Spooner, Eagle Genomics
- Audrey Kauffmann, Novartis

Two Surveys

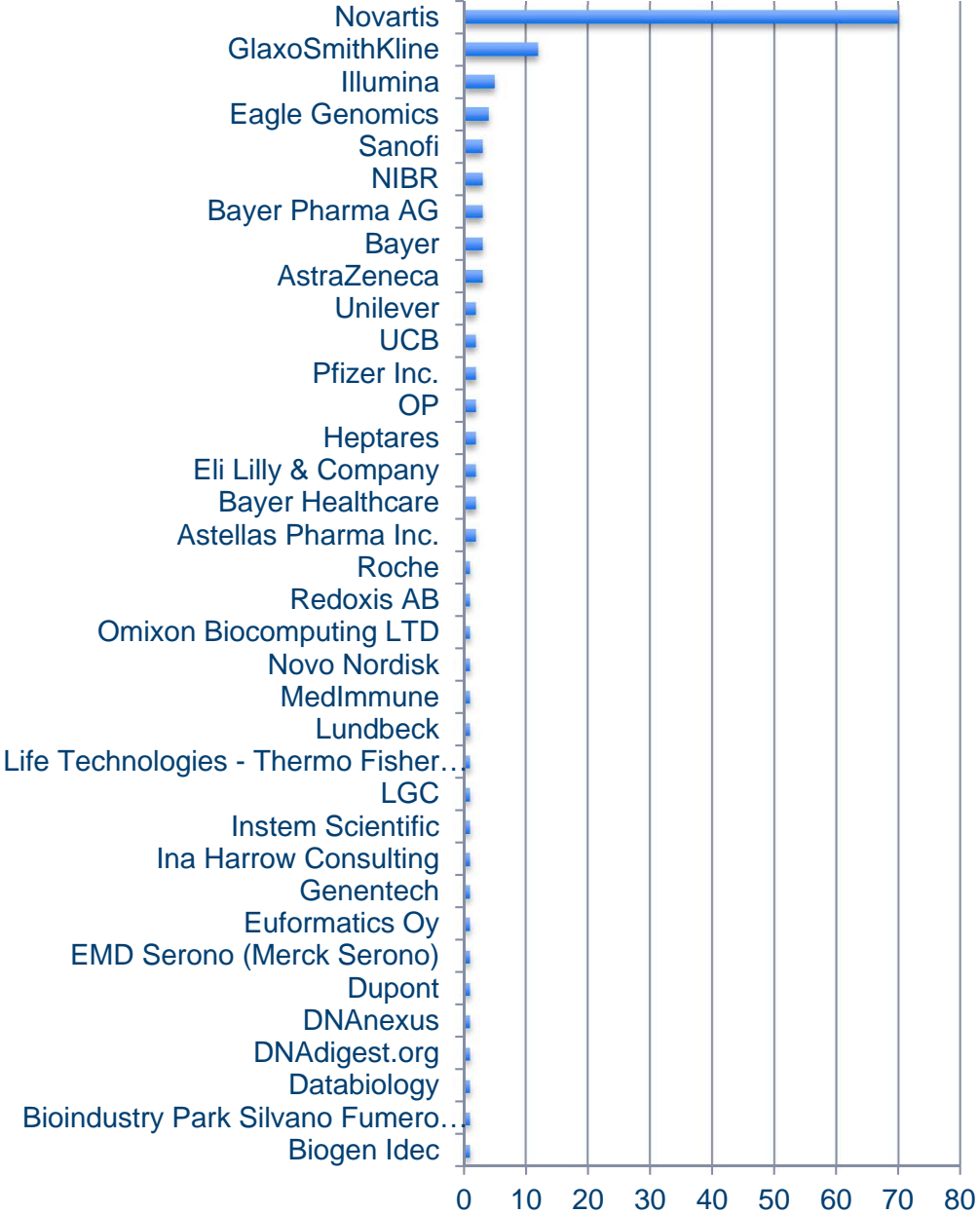
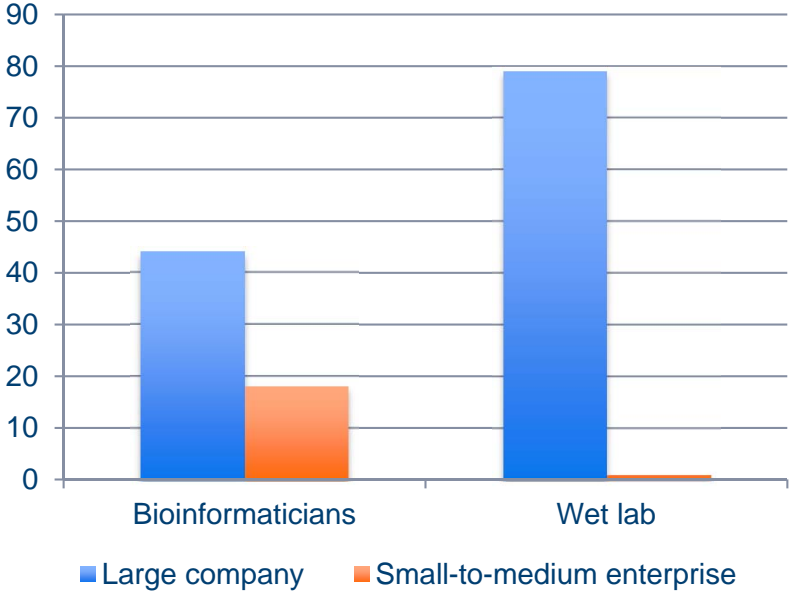


Life Science
Researchers

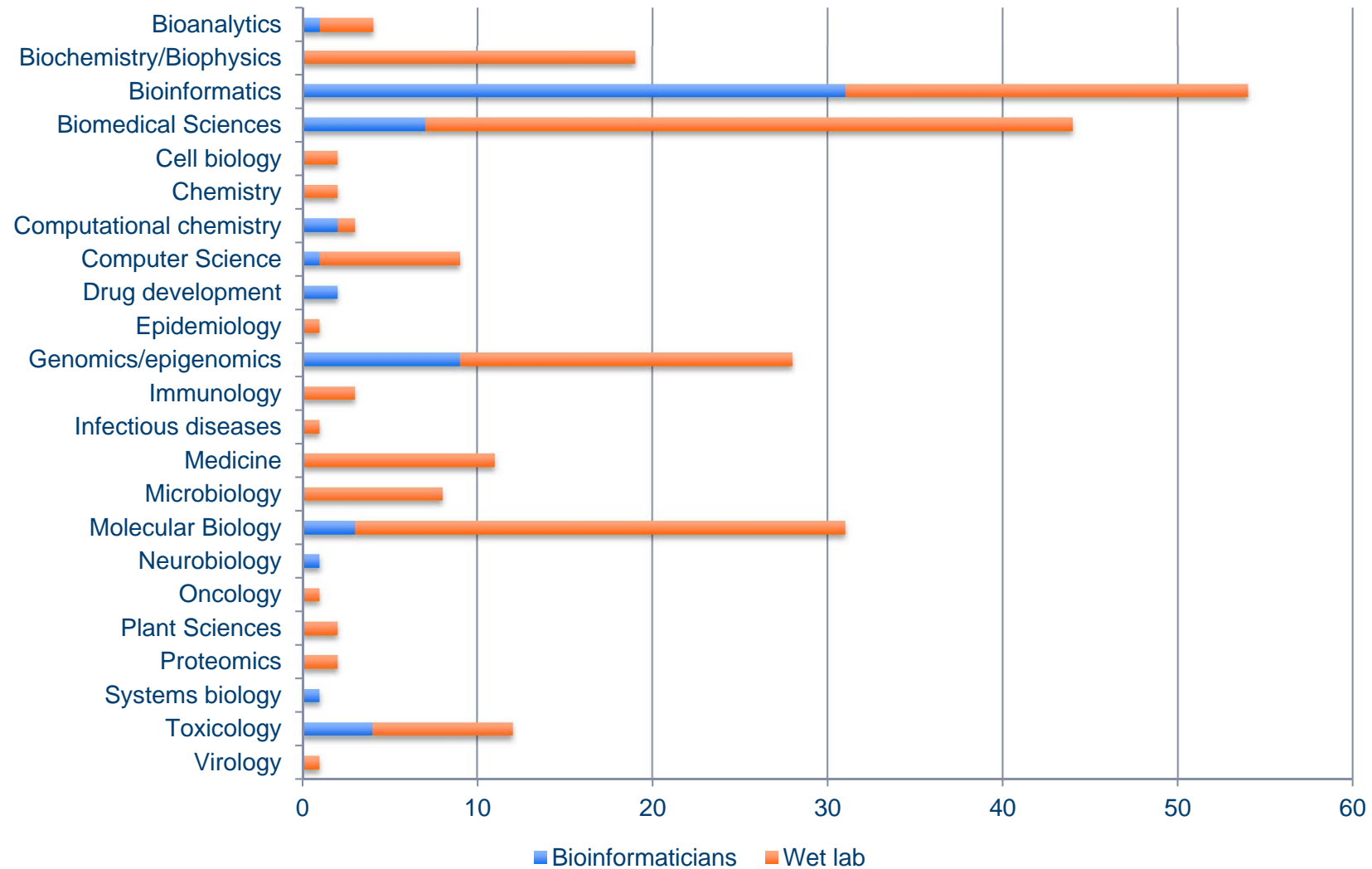
Technical
Infrastructure
Service Providers

- To help us understanding the bioinformatics-related training needs of industry and
- consequently to ensure that suitable training activities are developed and honed to target such needs.

Respondents

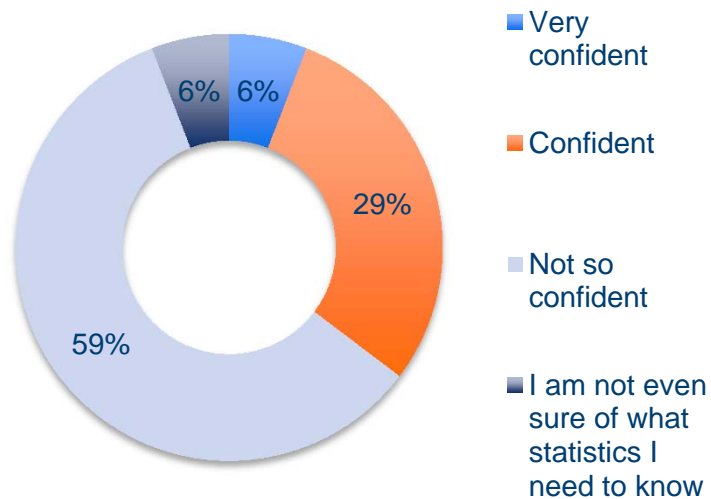


Disciplines



Lab-based scientists and statistics

How confident are you with statistics?



Do you collaborate with a bioinformatician/statistician?

Yes, I have a bioinformatician in the group that helps me to design experiments and also provides support for the data analysis

32.4%

Yes, occasionally I interact with a bioinformatician/statistician at my Institute, particularly when I get stuck and I don't know how to proceed.

32.4%

No, the data analysis is carried out by someone else. I just receive a file with the results.

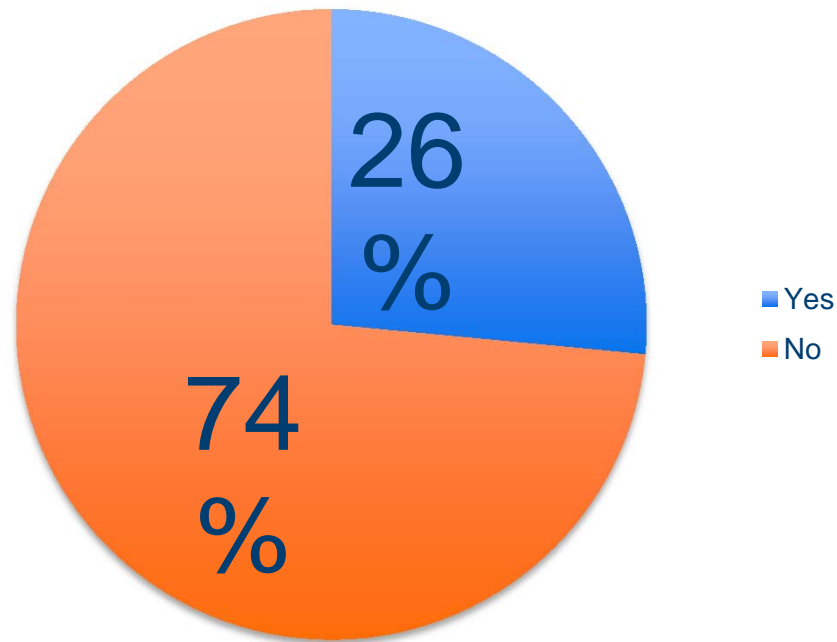
1.5%

No, I do not have any support. I am responsible for analyzing the data that I generate.

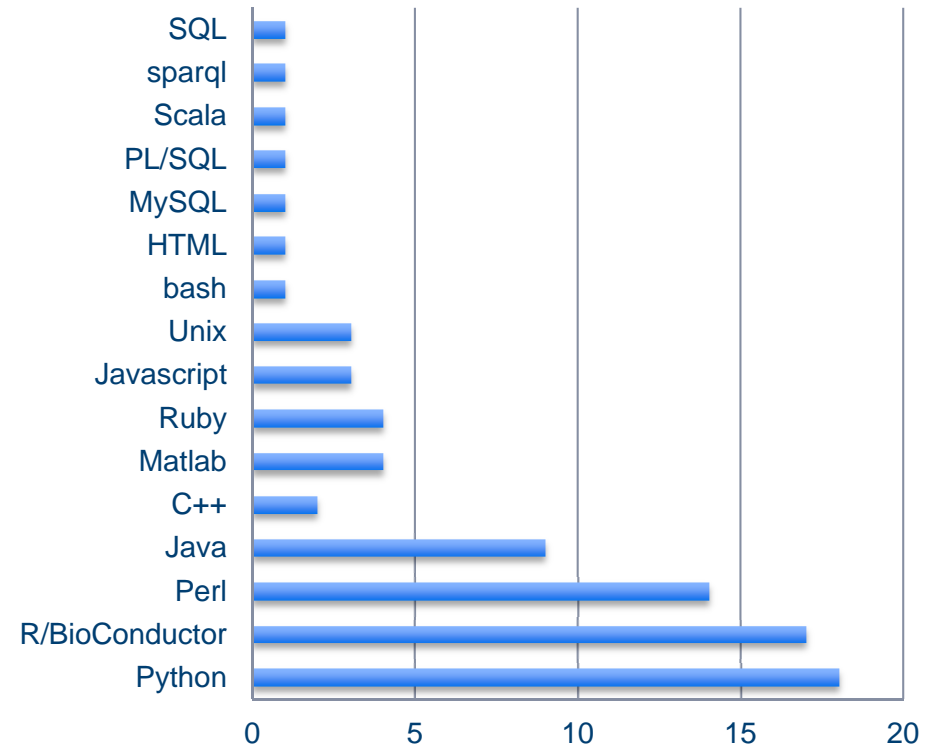
34.0%

Programming experience, languages

Wet lab – programming experience



Bioinformaticians - Programming languages

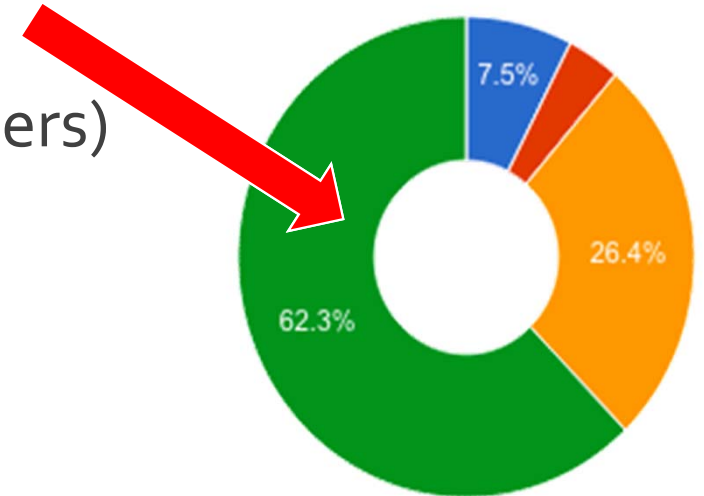


TRAINING STRATEGY

UK: How have we prioritised training need?

- Talking to research communities
- Surveys (both ours and others)
- Engaging with Industry
- Listening to experts (sector leads and others)
- Observing funding trends/initiatives

How confident are you with statistics?

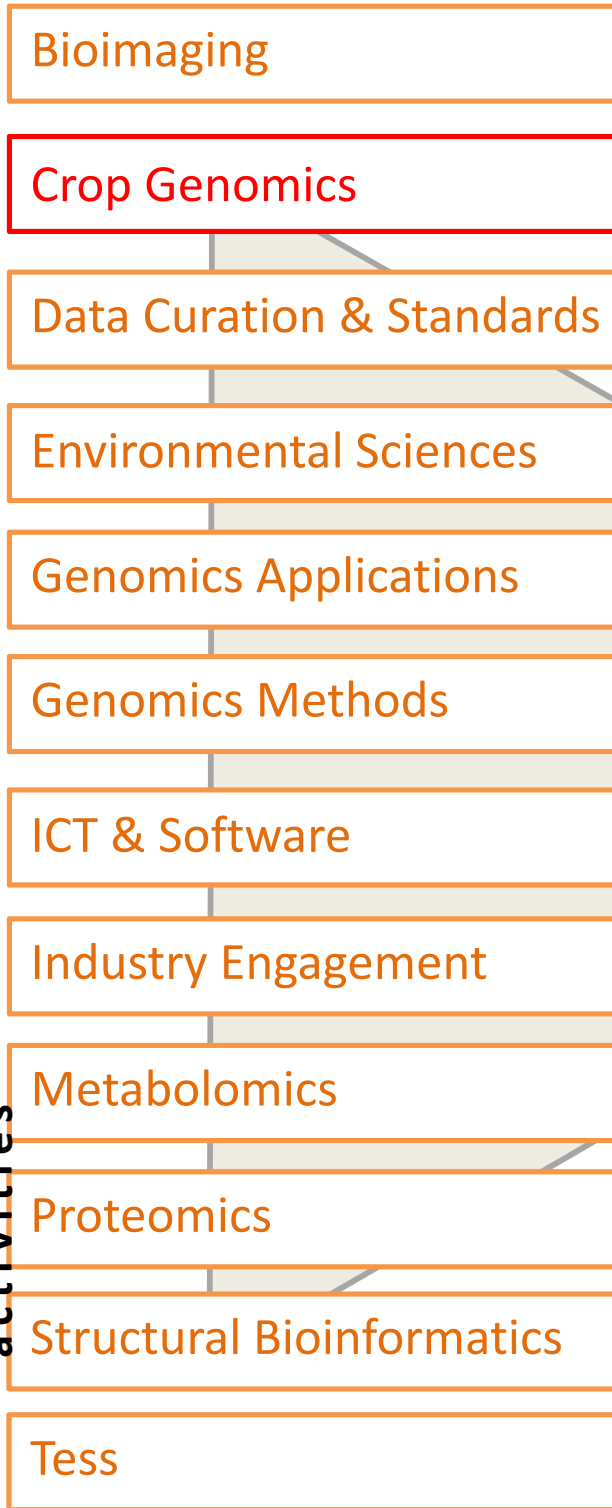


Confid...

Not so ...

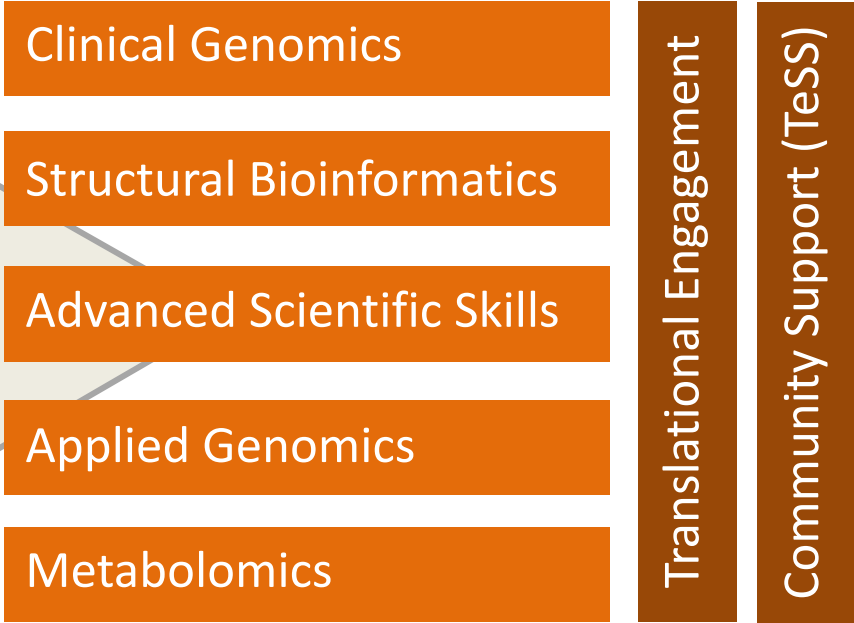
2/2

Current ELIXIR UK training sector activities

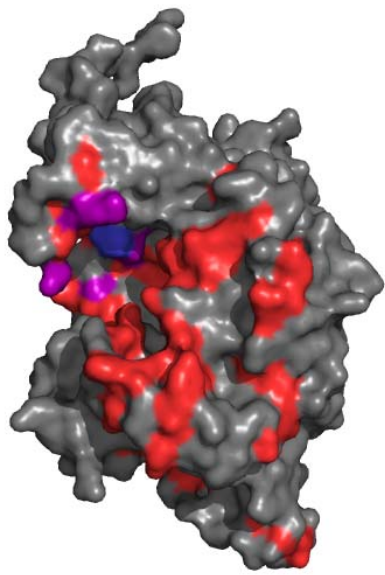


Statistics

Five areas to develop as ELIXIR UK strategic training priorities



Important supporting activities to develop further as ELIXIR UK



Structural Bioinformatics

**Genome
3D**

Protein structure classification/analysis

Alexey Murzin – LMB (Cambridge) - SCOP

Christine Orengo – UCL - CATH

Structural annotation of genome sequences and 3D models

Tom Blundell – Cambridge University - FUGUE

Christine Orengo – Gene3D

Julian Gough – Bristol University - SUPERFAMILY

David Jones – UCL - pDomThreader

Michael Sternberg – Imperial, London - PHYRE

- <http://genome3d.eu>

representing the
sector at initial
Training the Trainers
events

training gap analysis
and training
workflow workshop

fund-raising to plug
these training gaps.



Protein-protein interaction network analysis

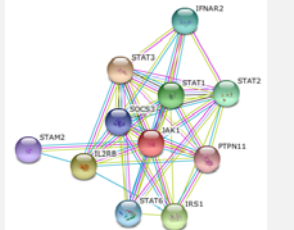
Database Integration

BioGRID 3.2 tutorial

IntAct tutorial

STRING 9.1 tutorial

STRING



Specific applications

DRUGBANK link

NCG4.0 Network of Cancer Genes link

3D structure

INstruct link

Interactome3D tutorial

Interactome3D

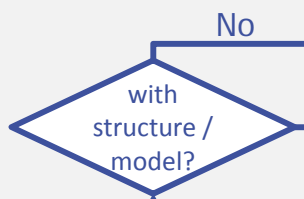


Visualisation & Functional annotation

Cytoscape tutorial

RCSB PDB
PROTEIN DATA BANK

BLAST



Structural analysis

Complex prediction

PRISM2.0 tutorial

Interface analysis

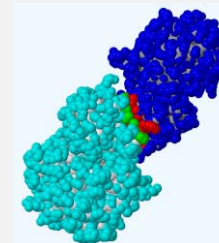
HOTREGION tutorial
A DATABASE OF COOPERATIVE HOTSPOTS

PDBePISA tutorial

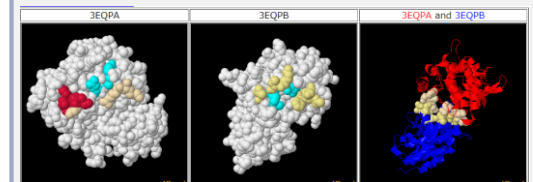
P.o.p.s comp link

Output

PDBePISA



HOTREGION



Genome 3D

*BLAST: Basic Local Alignment Search Tool (sequence similarity)

Identifying UK Training Needs

Bioimaging

Crop Genomics

Data Curation & Standards

Environmental Sciences

Genomics Applications

Clinical Genomics

ICT & Software

Industry Engagement

Metabolomics

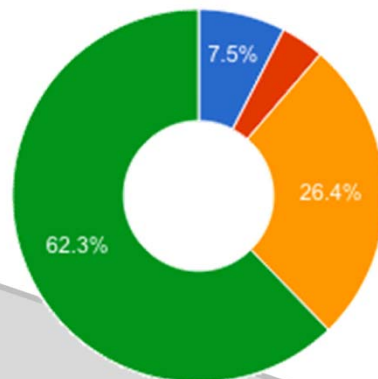
Proteomics

Structural Bioinformatics

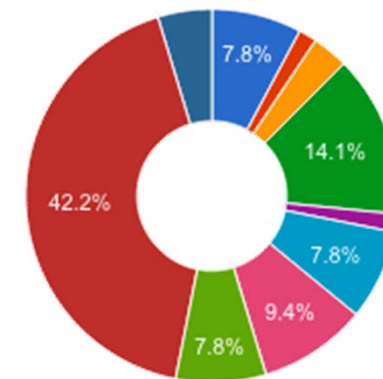
Tess

Industry & Sector-Specific Surveys

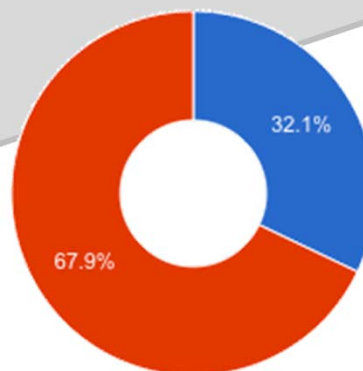
How confident are you with statistics?



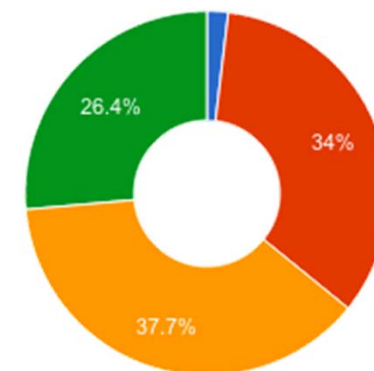
What is your job title/position?



Do you have any programming experience?



Do you collaborate with bioinformaticians?



Yes No

No, the... Yes, o... 1/2

TRAINING DELIVERY



Teach the “95% researchers” basic lab skills for scientific computing: the tools and techniques that will help them get more done in less time, and with less pain.
Volunteer instructors / Bootcamps / Train the trainers / Free lesson materials



www.software.ac.uk



software-carpentry.org

- Essential Software Engineering for researchers
- Software Sustainability Institute, UK
- UK and European workshops
 - Train researchers. Train the trainers.
 - Supporting other SW workshops
- Establishing SC Foundation
 - ELIXIR representation on interim board



DATA CARPENTRY

MAKING DATA SCIENCE MORE EFFICIENT

datacarpentry.org

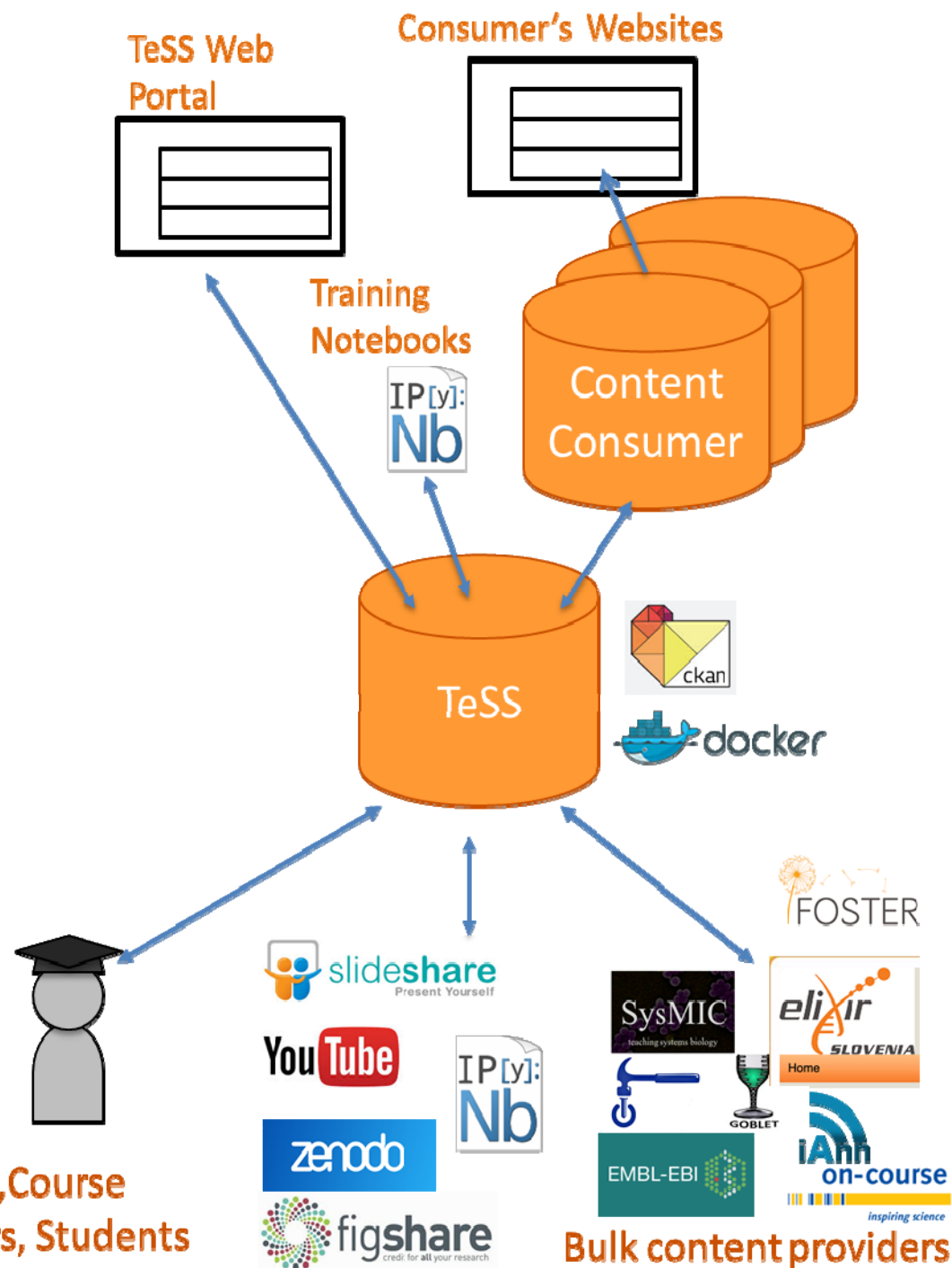


- Data literacy for researchers
- Expert data curation/integration
- Establishing Data Carpentry
 - ELIXIR representation on board
- First European Data Carpentry Workshop
Nov 27-28, 2014, UK
 - Applications to FOSTER open science training awards to scale up train the trainers

TeSS Portal

- Registering and discovering training materials
- Standard metadata
- Aggregated & Sourced from ELIXIR-UK, ELIXIR nodes and externals, Branding
- Packaging, VMs & linking
- Training workflows
- Progressively deliver forms of training online
- Cooperation with eLearning Platform, **ELIXIR-Slovenia**
- Piloting with Structural Biology

Lecturers, Course
Organizers, Students



Bulk content providers

Summary



NETHERLANDS

SWITZERLAND

DENMARK

EBI

SWEDEN

SLOVENIA

ESTONIA

ITALY

Training

Cloud
Technical
Services



Data
Interoperability,
vocabularies and
ontology
services

Tools
Interoperability
& Service
Registry



Thank you



<http://elixir-uk.org/>

Summary



NETHERLANDS

SWITZERLAND

DENMARK

FRANCE

SWEDEN

SLOVENIA

Training

Cloud
Technical
Services



Data
Interoperability
vocabularies and
ontology
services

Tools
Interoperability
& Service
Registry



EDAM, SWO

Questions from the floor

- How many people will be trained, or can we aim to train? Scalability and multipliers
 - (the model of train the trainers, coordinated materials and bootcamps is a scalable approach. The addressing of scale of training was appreciated.)
- How do we relate to EMTRAIN, Coursera, ROSALIND (<http://rosalind.info>)
 - (we feed from and to these as resources for TeSS, but we need more formal links).
- How will we measure the impact of the training?
 - (follow-up metrics needed. Hard problem. Manny in a metrics TF).
- How will our training be applied to clinical and medical training, esp in different ELIXIR countries with their own governance and healthcare regulations.
 - (serious local customisation needed)
- How do we get help to run a SC bootcamp in our node?
 - (we need to have core information on the web site).

Additional thoughts

- We should consider the ELIXIR-UK activities in terms of Additional and Commissioned Services, in line with the Consortium Agreement.

ELIXIR UK node

Mission: To facilitate training of research scientists and infrastructure technologists in bioinformatics, computing, statistics and biology, in partnership with UK centres, industry and other ELIXIR Nodes.



2014

Q1: Recruitment round 1 & ELIXIR PoW5 Training drafted & ELIXIR Training Node Survey & Structural Bioinformatics Training Workshop & Industry workshop @ TGAC

Q2: ELIXIR HoN and TCG meeting @Sandhamn & Lee, Aleks, Niall, Milo start & Advisory Group formed & TeSS meeting Manchester

Q3: Industry Committee constituted, & Kick-off meeting, Rita starts, ELIXIR/ELIXIR-UK Node meeting Hinxton, TrCC constituted

Q4: ELIXIR Innovation and SME forum & ELIXIR SAB - Heads of Nodes meeting, Hinxton, ELIXIR-UK ECA being negotiated

INFRADDEV-3 preparation

Strategic Focus on:
Clinical Genomics
Structural Bioinformatics
Advanced Scientific Skills
Applied Genomics
Metabolomics

Key achievements in Software & Data Carpentry delivery and International engagement (BD2K, SCF, RDA etc)

Bringing together an ELIXIR-wide Training Coordination Group

Oxford University Computational Genomics Analysis and Training (CGAT)
University of Manchester
European Bioinformatics Institute (EMBL-EBI)
University of Cardiff & NERC EOS Centre
The Genome Analysis Centre (TGAC)
The Oxford e-Research Centre
University College London
University of Birmingham
University of Edinburgh
Queen Mary, London
University of Cambridge
University of Liverpool Centre for Genomic Medicine

Activities

Training

Research scientists and
Infrastructure technologists
bioinformatics, computing, statistics
and biology, service deployment

Multiplier effects

Partnerships

12 UK centres, National Centres

Software Sustainability Institute, Digital Curation Centre.

ELIXIR Nodes

SIB, NL, Slovenia, Sweden, EBI

International initiatives

GOBLET, Software Carpentry, Data Carpentry, FOSTER, Intl Soc

BioCuration

Industry



Activities

Research scientists and Infrastructure technologists
bioinformatics, computing, statistics and biology,
service deployment
Multiplier effects

Training

Data
Interoperability,
vocabularies and
ontology services

Cloud
Technical
Services



 Windows Azure
€32,000

Tools
Interoperability
& Service
Registry




Activities

Research scientists and Infrastructure technologists
bioinformatics, computing, statistics and biology,
service deployment
Multiplier effects



Training



Data
Interoperability,
vocabularies and
ontology services

NETHERLANDS
Bring Your Own Data



Cloud
Technical
Services



 Windows Azure
€32,000

ontology services



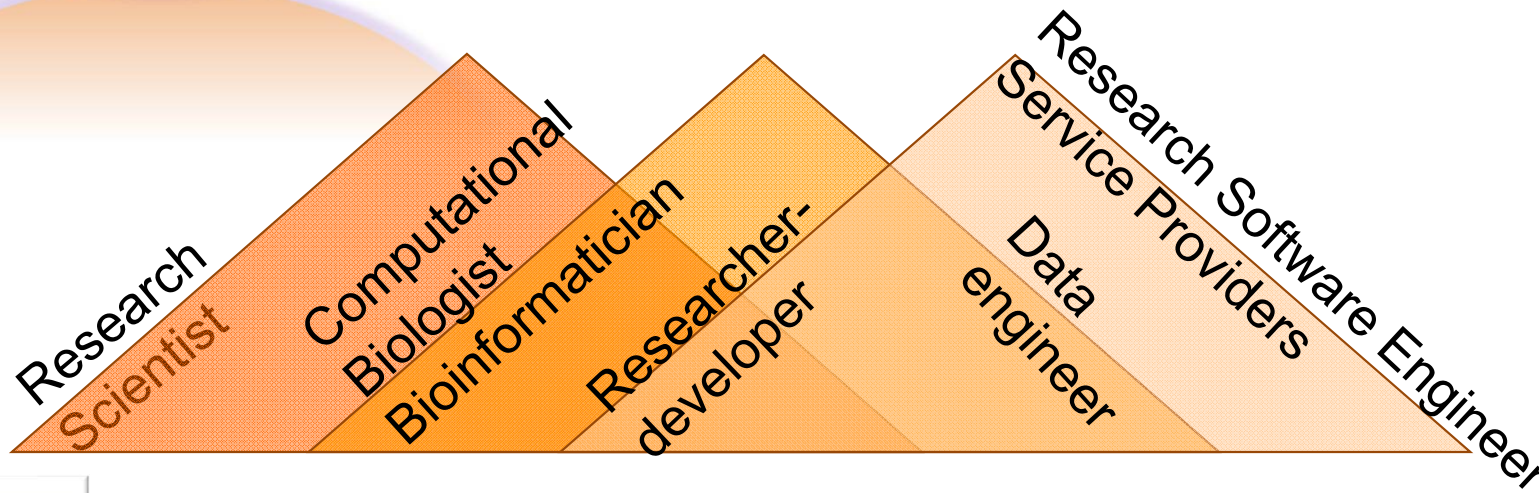
Tools
Interoperability
& Service
Registry

DENMARK
SWEDEN
EBI

registry

Training

Research scientists and Infrastructure technologists
bioinformatics, computing, statistics and biology, service deployment



Genomic Research Researchers

Lee Larcombe



Technical Infrastructure Infostructure Providers

Aleks Pawlik





Online Training

A number of scripts, examples etc are available:

Statistics and R

- [Introduction to R](#)
- [Data input and simple plots](#)
- [Heatmaps in R](#)

Microarray analysis

- [Working with Affymetrix CEL files in R](#)

NGS Read Mapping

NGS De novo assembly

NGS Metagenomics

- [A quick and dirty analysis of some human microbiome data](#)

Analysis of microRNA-Seq data

- [Counting known microRNAs in five easy steps](#)

Genomics

- [How big are pig exons? \(uses R\)](#)
- [How much data will my sequencers produce?](#)

Fun Stuff

- [R for geeks on Valentines Day](#)

Project Enquiries

Grassroots Genomics

TGAC Fellowships

Training Events

[Software Carpentry: Instructor Training](#)

[Training School: genotyping by sequencing](#)

[Bioinformatics WikiBook Collaborative Workshop](#)

[TGAC Winter School: Patterns, similarities and differences - from bioinformatics to biology](#)

[Statistics for Ecology, Genetics and Genomics using R](#)

Events

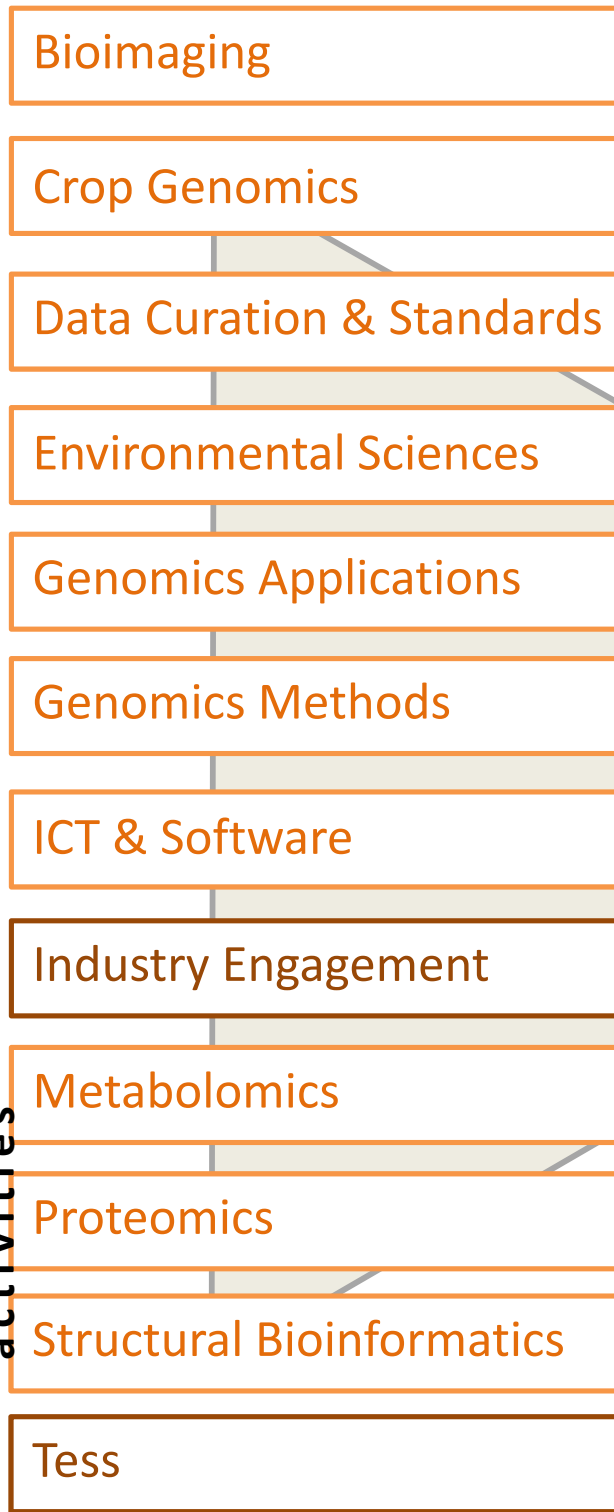
▾ Online Training

- [Counting known microRNAs in five easy steps](#)
- ▶ [EU Training Course](#)
- [Heatmaps in R](#)
- [How big are pig exons?](#)
- [Introduction to R](#)
- [Next-gen output](#)
- [Quick and dirty analysis of human microbiome data](#)
- [R for geeks on Valentines day](#)
- [Working with Affymetrix CEL files](#)

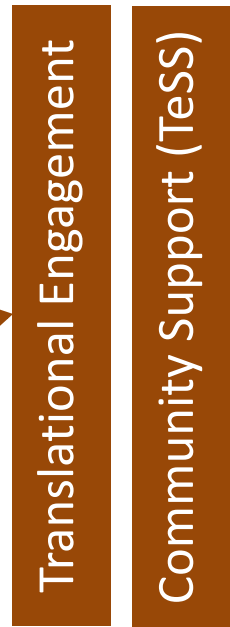
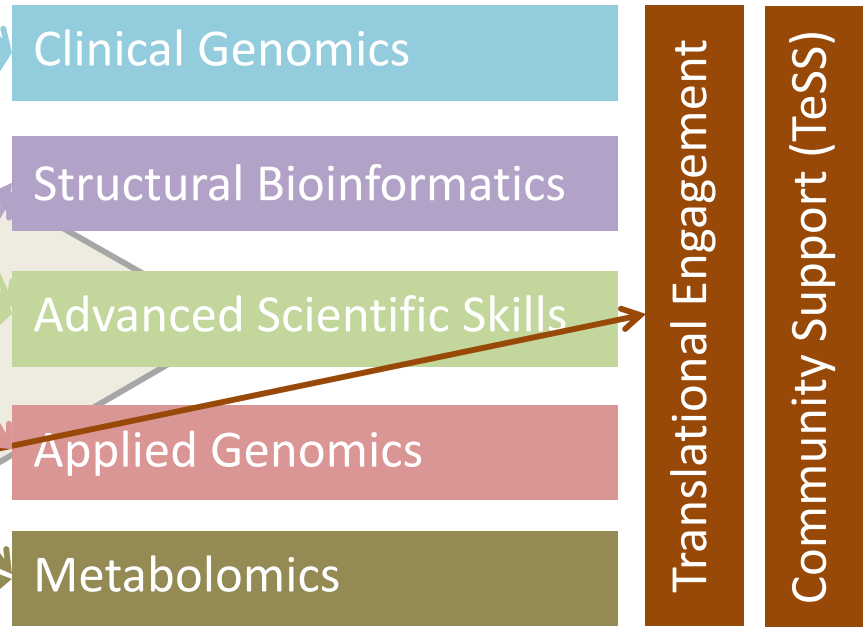


- 22&23 October 2014
 - ELIXIR UK & SSI **Software Carpentry “Train the Trainer”** Workshop @TGAC –ELIXIR UK
- 6&7 November 2014
 - Advanced R" Software Carpentry bootcamp at UZH/ETHZ in Zurich
- 9&14 November 2014
 - Software Carpentry workshop for **NHS Clinical Bioinformatics MSc** @ Manchester
- 17&18 November 2014
 - CGR & GARNet Software Carpentry for the 2nd Plant Scientists Bootcamp @Liverpool
- 27&28 November 2014
 - ELIXIR UK **Data Carpentry workshop** in Manchester (data curation & standards and tools module)
- December 2014/ January 2015
 - ELIXIR UK **Microsoft Azure for Research** workshop in Manchester
- 3-4 February 2015
 - ELIXIR UK & SSI Software Carpentry Course for **EPSRC & MRC Centre for Doctoral Training in Regenerative Medicine**

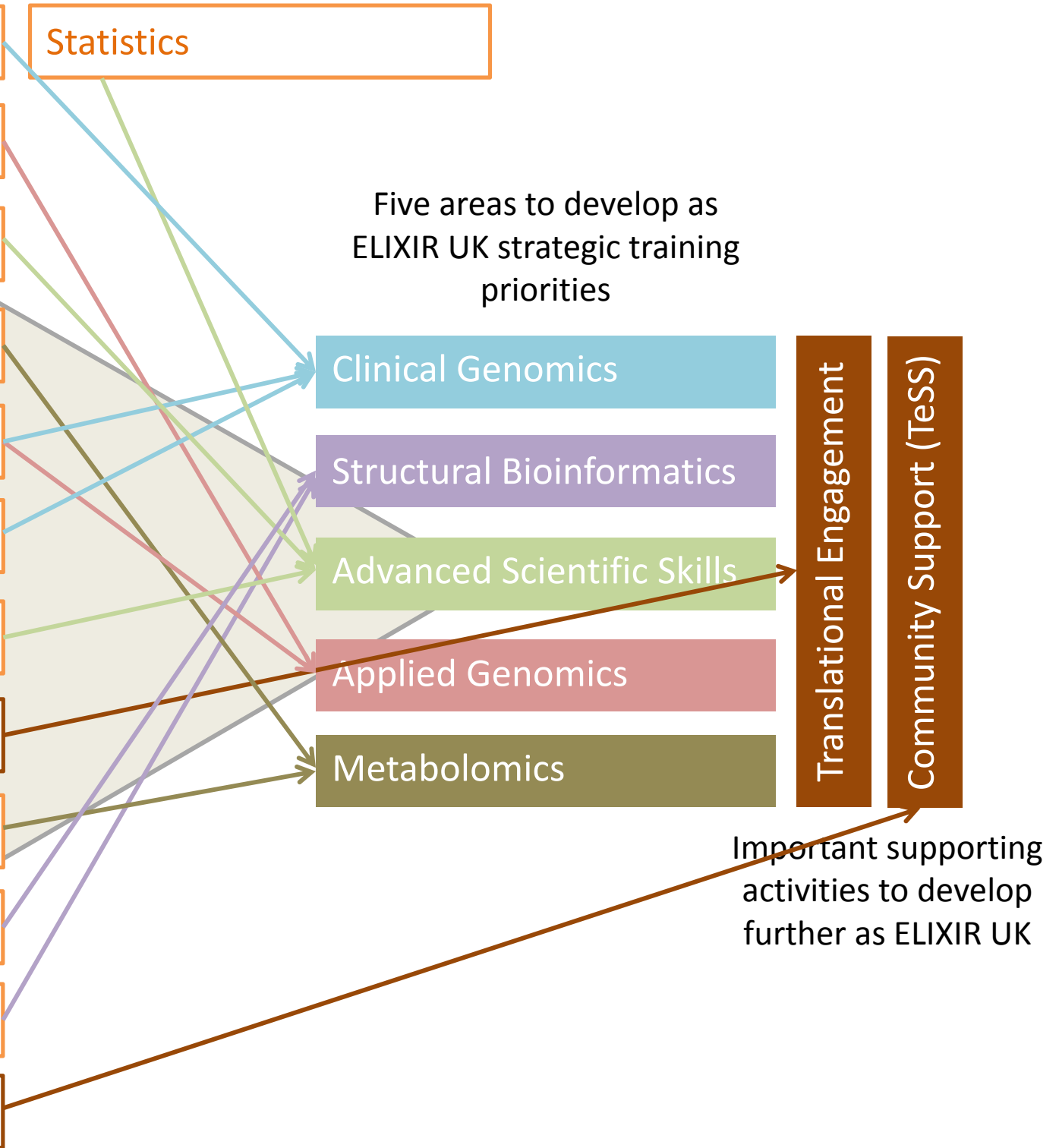
Current ELIXIR UK training sector activities



Five areas to develop as ELIXIR UK strategic training priorities

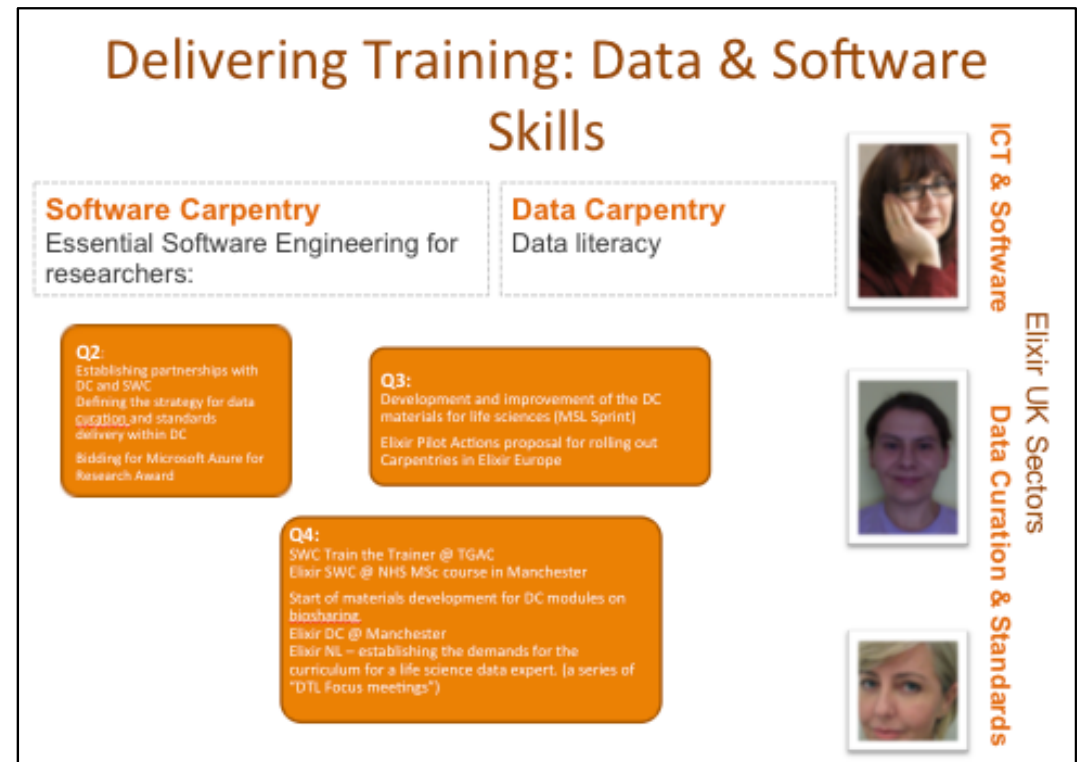


Important supporting activities to develop further as ELIXIR UK



Building International Reach

- Software carpentry
- Data carpentry
- SCF
- NIH BD2K initiatives



Current training achievements in these sectors will feed into Advanced Scientific Skills – becoming a model for international outreach in other priority areas

ELIXIR-UK working across Europe: Training Coordination Committee

Establish an interacting ELIXIR wide training community & to ensure coherency in the delivery of training related to ELIXIR activities.

TrCC members:

BE	Katrijn Vannerum
CZ	Daniel Svozil
DK	Peter Longreen
EE	Hedi Peterson
FI	Eija Korpelainen
FR	Julie Thompson
IL	Michal Linial
IT	Allegra Via
NL	Celia van Gelder
NO	Ståle Nygård
PT	Pedro Fernandes
SI	Brane L. Leskosek & Peter Juvan
ES	Oswaldo Trelles
SE	Sara Light
CH	Patricia Palagi
UK	Rita Hendricusdottir & Lee Larcombe
EMBL-EBI	Sarah Morgan
TeSS	Carole Goble (invited for INFRADEV3)

Chair: Rita Hendricusdottir

