Reproducibility and standards: ELIXIR Training Platform perspective

Patricia Palagi, co-leader ELIXIR Training Platform

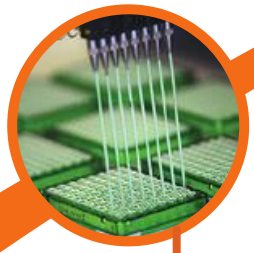
SIB Swiss Institute of Bioinformatics, ELIXIR Switzerland

www.elixir-europe.org

ELIXIR connects national bioinformatics centres and EMBL-EBI into a sustainable European infrastructure for biological research data

medicine

agriculture



bioindustries

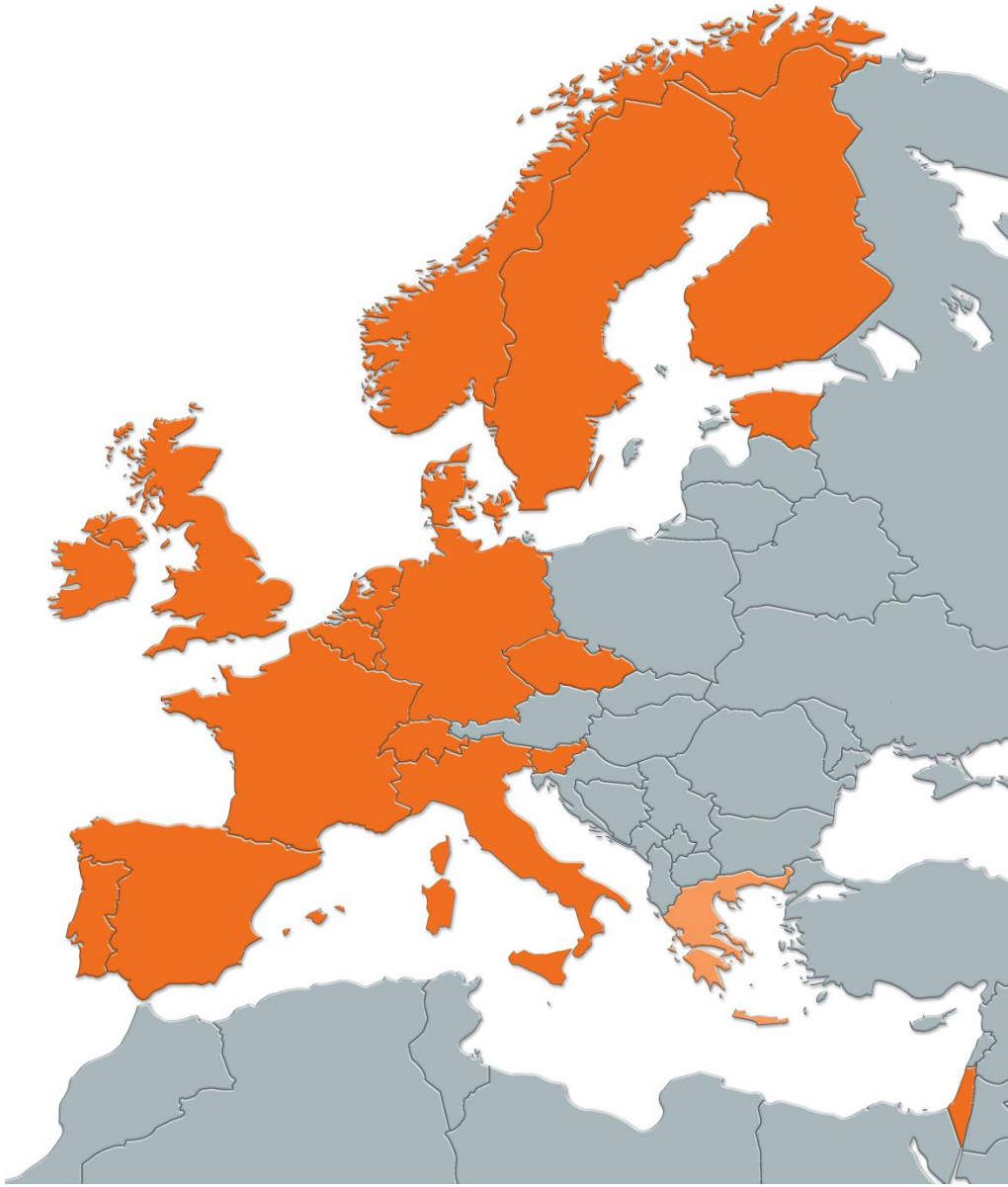
environment

ELIXIR underpins life science research – across academia and industry

ELIXIR Membership

ELIXIR Members

ELIXIR Observers



ELIXIR Structure

Five technical platforms for Data, Tools, Interoperability, Compute and Training

Complemented by four Use Cases for marine meta-genomics, rare diseases, human data and plants sciences

Training: Professional skills for managing and exploiting data

- Mission: provide developers, researchers and trainers with skills to use and exploit ELIXIR services
- Builds on and complements trainings in ELIXIR Nodes

Chris Ponting, UK, **Patricia Palagi,** CH, **Celia van Gelder,** NL, **Rita Hendricusdottir,** UK

Training: Professional skills for managing and exploiting data

- ELIXIR Training infrastructure:
  - TeSS portal to share and exchange training content and events: http://tess.elixir-uk.org
  - E-learning platform and solutions
  - Large pool of ELIXIR trainers (Train the Trainer)
  - Good training practices and guidelines, including metrics and evaluation
  - Training events for researchers, developers, trainers and infrastructure operators

We need standards to

Describe Training related:

- Measures and metrics
- Events
- Materials
- Trainer competencies

Standardised measures - training quality, success & impact

1st step: Survey the community – identified descriptors in 3 categories

- descriptors for training programs
- quantitative metrics for individual training events
- questions for short-term feedback

Standardised measures - training quality, success & impact

1st step results - Some descriptors

- **quantitative data** - course program:
  - #events, #days, #participants, geographical spread
- **trainers' quantitative data** - effort involved:
  - #trainers, #hours in preparation/delivery/post course
- **course ratings** - overall organization, course content, balance and level, evaluations of each day of the course
- **qualitative data** - capturing suggestions for improvements and additional training topics.

Standardised measures - training quality, success & impact

Next steps – descriptors of impact on

- **productivity and quality** – publications, citations, skills learned useful to the research, to validate results, data analysis shorter
- **career trajectory of the trainee** – hired, promoted, same job but accomplished better, new collaborations
- **career trajectory of the trainer** – efforts acknowledged or validated in job, hired, promoted, same job but accomplished better, new collaborations

Standardising the process

And then next steps:

- From the set of common descriptors – **create a standard**
- Define a mechanism to ensure and facilitate capturing this information from Nodes
- Best practices and guidelines on standard and capturing
- Seek commitment of the community to adopt the standard

TCC BD2K Training Coordinating Center

NIH Big Data to Knowledge(BD2K)

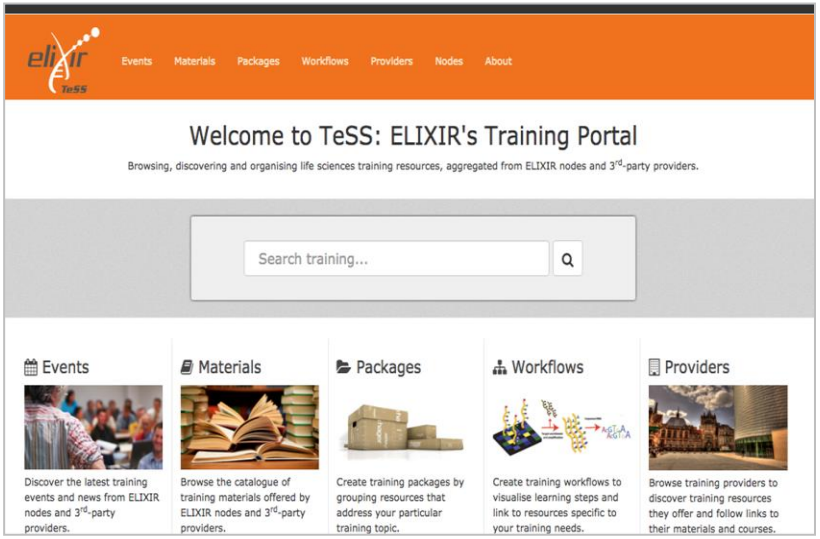
GOBLET Global Organisation for Bioinformatics Learning, Education & Training

elixir

ELIXIR TeSS Portal (Training eSupport System)

for browsing, discovering and organizing training events, courses and materials

https://tess.elixir-uk.org/



ELIXIR TeSS Portal (Training eSupport System)

https://tess.elixir-uk.org/

Aims to give a snapshot of the ELIXIR training landscape

- automatically aggregating data from Nodes & 3rd-party providers
- making events & resources discoverable through useful filtering tools
- surfacing information to support user decisions & choices via training packages & workflows
- forming a resource network by linking data with other ELIXIR registries such as bio.tools and biosharing

Allows Nodes to contribute their training resources

- showcase latest news, events, activity highlights, *etc*.

TeSS driving development of standard for sharing training materials and events

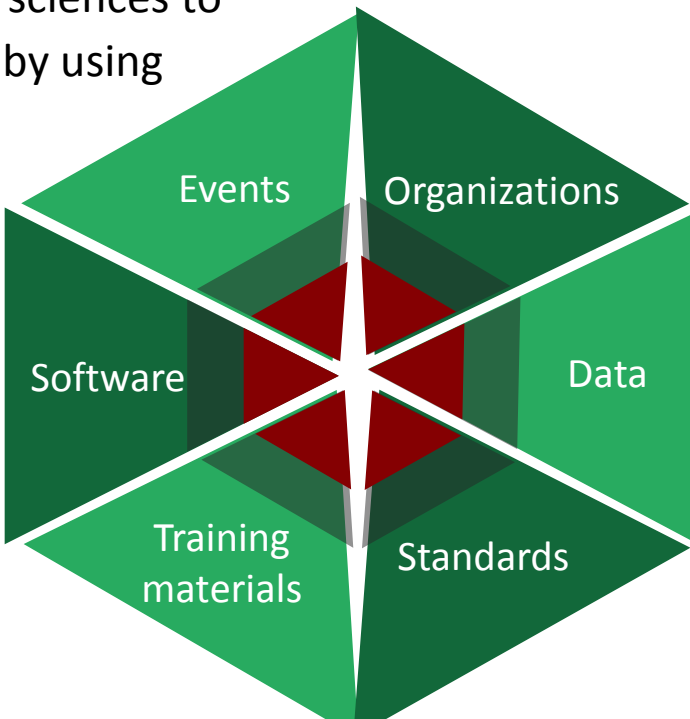
Bioschemas

- improves data discoverability in life sciences by improving the exposure of our data repositories:
  - By search engines like Google
  - By aggregated portals like TeSS
- It does this by encouraging content providers in life sciences to expose consistent structured data in their websites by using schema.org mark-up.

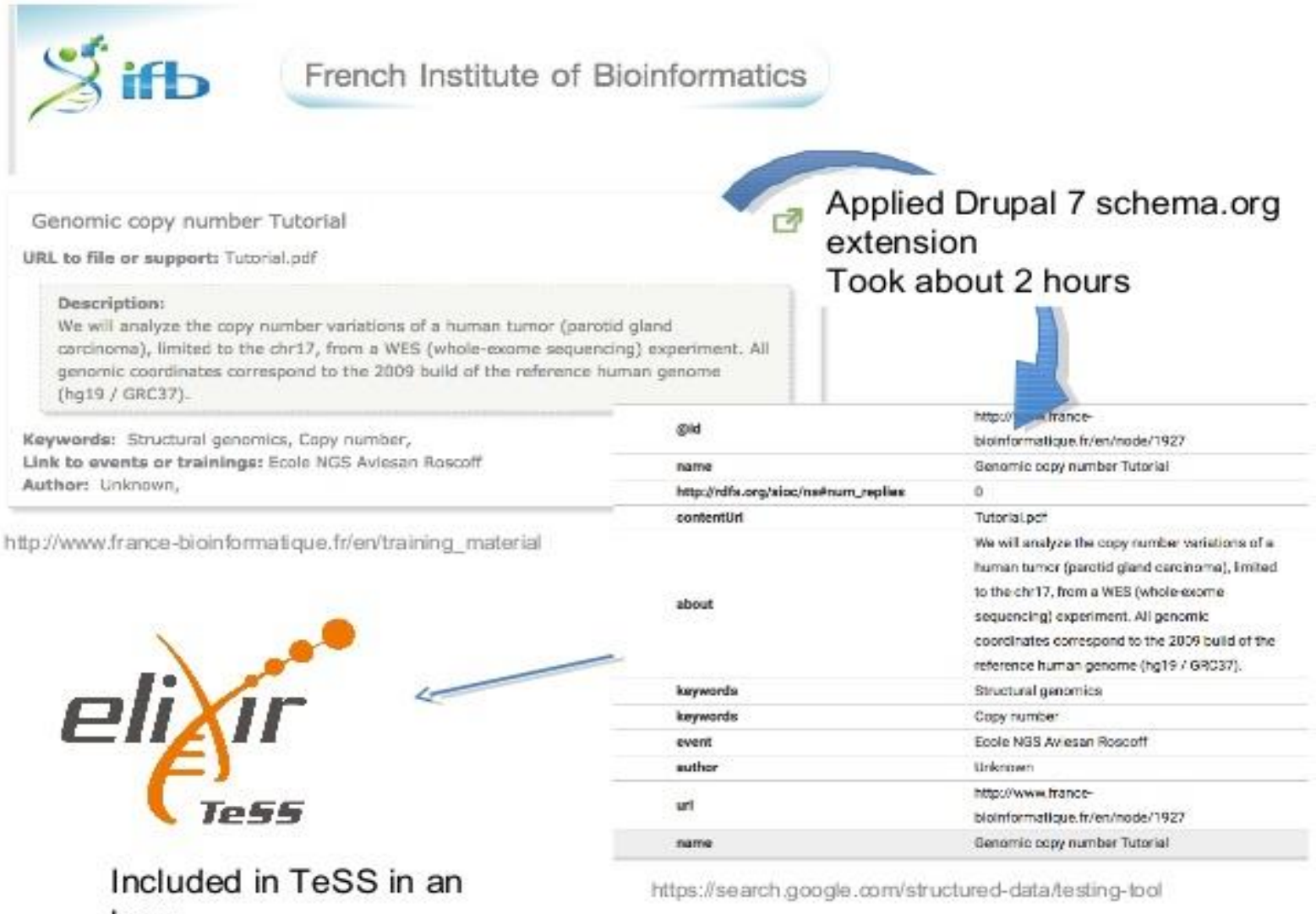
In TeSS

- Describes all training events and materials using Bioschemas
- Parses, aggregates and integrates any Bioschemas compliant training resource metadata

Identifier, Title, Description, Author, Topics, Audience, Publication Date, …

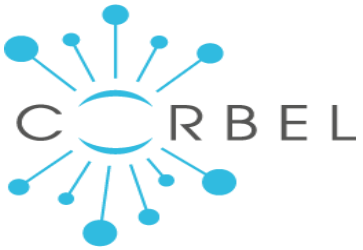


Bioschemas compliant metadata integration into TeSS



Competency profiles

- Standards to describe the knowledge, skills, values and attitudes necessary for a profession or a particular job role
- Initiated in ISCB
  - developing competency profiles for bioinformatics users, scientists and engineers. Examples: Welch, L. et al. (2014) *PLoS Comp. Biol.*
- Applying such profiles to define professionals and the identify training needs



We care for reproducibility

In Training related:

- Courses
- Materials

Insuring courses can be reproduced

Course materials (slides, exercises, datasets)

- Training materials annotated, findable in TeSS and reusable
- Hackathons: Metagenomics, Data carpentry, etc.

Clouds, virtual machine (VM) images and Docker

- Overcoming technical problems during courses
- Containers pre-installed
- Cloud platforms: more computing power and memory, scaling for running hundreds of simultaneous jobs

Distance learning

- Enable remote execution of a live course: teacher in one location and students on remote and distributed locations

Train the trainers - why

- Need for courses outweighs the number of places available, notably in NGS courses
- Life sciences expand into newer territories
- Need for bioinformatics competencies (and hence training):
  - well recognised
  - ability to provide such training - not yet well developed in all ELIXIR Nodes

New trainers are therefore required across both academia and industry

Train the trainers

Build a highly skilled and coherent community of trainers

- Tools and tips for providing an enriching learning experience to their trainees
- Guidance on course development
- Access to a wider support network, with a focus on ELIXIR-derived resources and infrastructure

Development

- Reproducible framework and associated material
- Guidelines for delivery of TtT courses

GOBLET
Global Organisation for Bioinformatics Learning, Education & Training

elixir

Training for Standards and Reproducibility

We are responsible for:

Scientists
Developers
Trainers

ELIXIR community and beyond

Courses - Data management and data stewardship

- Data-intensive analyses: data-management expertise is essential
- Scientists need to ensure that:
  - their experiments are reproducible
  - the resulting data ("data stewardship") are :
    - accessible
    - reusable
    - sustainable

- ELIXIR priority: providing solutions, best practices and training

ELIXIR Software and Data Carpentry programme

- Bring the model and format of Carpentry training to ELIXIR and its Nodes
- Several workshops organised within Nodes
- **~ 240 life science researchers and instructors trained**
- First ELIXIR instructors received their certificates in February 2016
- Many more yet to come in collaboration with Software and Data Carpentry

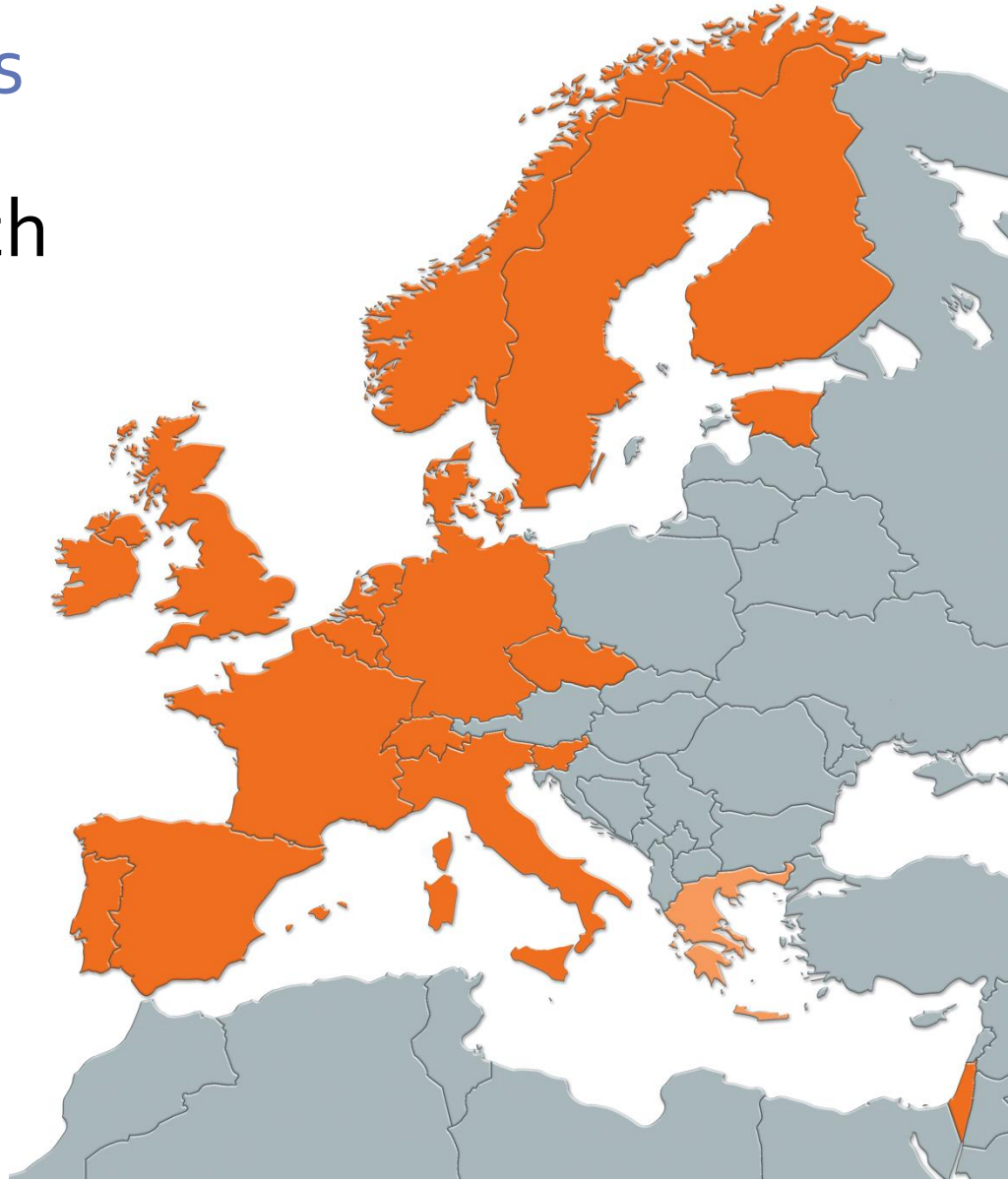
software carpentry

DATA CARPENTRY
MAKING DATA SCIENCE MORE EFFICIENT

elixir

Training in ELIXIR Nodes

- Reproducible research
- Experimental design
- Statistical analysis



Acknowledgements

Aleksandra Nenadic, Alex Botzki, Allegra Via, Björn Grüning, Brane Leskošek, Carole Goble, Cath Brooksbank, Celia van Gelder, Chris Ponting, Eija Korpelainen, Finn Bacall, Frederik Coppens, Gabriella Rustici, Hedi Peterson, Heinz Stockinger, Henriette Husum Bak-Jensen, Jessica Lindvall, Jure Dimec, Lee Larcombe, Louisa Bellis, Milo Thurston, Niall Beard, Oscar Torreno, Pedro Fernandes, Rafael C Jimenez, Rita Hendricusdottir, Salvador Capella Gutierrez, Sarah Morgan, Ståle Nygård, Susanna-Assunta Sansone, Terri Attwood, Vera Matser, Victoria Dominguez del Angel, Vojtech Spiwok, Wei GU

Acknowledgements – Quality/Success/Impact

Aleksandra Nenadic, Alex Botzki, Allegra Via, Björn Grüning, Brane Leskošek, Carole Goble, Cath Brooksbank, Celia van Gelder, Chris Ponting, Eija Korpelainen, Finn Bacall, Frederik Coppens, **Gabriella Rustici**, Hedi Peterson, Heinz Stockinger, Henriette Husum Bak-Jensen, Jessica Lindvall, Jure Dimec, **Lee Larcombe**, **Louisa Bellis**, Milo Thurston, Niall Beard, Oscar Torreno, Pedro Fernandes, Rafael C Jimenez, Rita Hendricusdottir, Salvador Capella Gutierrez, **Sarah Morgan**, Ståle Nygård, Susanna-Assunta Sansone, Terri Attwood, Vera Matser, Victoria Dominguez del Angel, Vojtech Spiwok, Wei GU

Acknowledgements - TeSS

Aleksandra Nenadic, Alex Botzki, Allegra Via, Björn Grüning, Brane Leskošek, **Carole Goble**, Cath Brooksbank, Celia van Gelder, Chris Ponting, Eija Korpelainen, **Finn Bacall**, Frederik Coppens, Gabriella Rustici, Hedi Peterson, Heinz Stockinger, Henriette Husum Bak-Jensen, Jessica Lindvall, Jure Dimec, Lee Larcombe, Louisa Bellis, **Milo Thurston**, **Niall Beard**, Oscar Torreno, Pedro Fernandes, Rafael C Jimenez, Rita Hendricusdottir, Salvador Capella Gutierrez, Sarah Morgan, Ståle Nygård, **Susanna-Assunta Sansone**, **Terri Attwood**, Vera Matser, Victoria Dominguez del Angel, Vojtech Spiwok, Wei GU

Acknowledgements – Train the trainers

Aleksandra Nenadic, Alex Botzki, **Allegra Via**, Björn Grüning, Brane Leskošek, Carole Goble, Cath Brooksbank, Celia van Gelder, Chris Ponting, Eija Korpelainen, Finn Bacall, Frederik Coppens, Gabriella Rustici, Hedi Peterson, Heinz Stockinger, Henriette Husum Bak-Jensen, Jessica Lindvall, Jure Dimec, Lee Larcombe, Louisa Bellis, Milo Thurston, Niall Beard, Oscar Torreno, Pedro Fernandes, Rafael C Jimenez, Rita Hendricusdottir, Salvador Capella Gutierrez, **Sarah Morgan**, Ståle Nygård, Susanna-Assunta Sansone, Terri Attwood, Vera Matser, Victoria Dominguez del Angel, Vojtech Spiwok, Wei GU

Acknowledgements

Aleksandra Nenadic, Alex Botzki, Allegra Via, Björn Grüning, Brane Leskošek, Carole Goble, Cath Brooksbank, Celia van Gelder, Chris Ponting, Eija Korpelainen, Finn Bacall, Frederik Coppens, Gabriella Rustici, Hedi Peterson, Heinz Stockinger, Henriette Husum Bak-Jensen, Jessica Lindvall, Jure Dimec, Lee Larcombe, Louisa Bellis, Milo Thurston, Niall Beard, Oscar Torreno, Pedro Fernandes, Rafael C Jimenez, Rita Hendricusdottir, Salvador Capella Gutierrez, Sarah Morgan, Ståle Nygård, Susanna-Assunta Sansone, Terri Attwood, Vera Matser, Victoria Dominguez del Angel, Vojtech Spiwok, Wei GU

And many more, that I have certainly forgotten and I apologise in advance ….

Acknowledgements

Over 160 institutes involved in ELIXIR

www.elixir-europe.org

[@ELIXIREurope](https://twitter.com/ELIXIREurope)

[/company/elixir-europe](https://www.linkedin.com/company/elixir-europe)

SIB Swiss Institute of Bioinformatics

elixir SWITZERLAND elixir