Data stewardship training resources in ELIXIR

Celia van Gelder (DTL/ELIXIR-NL)

Open Science and Research Data Management Train-the-Trainer Bootcamp
April 17, 2019

www.elixir-europe.org

Content

ELIXIR & ELIXIR Training Platform

- Intro
- Highlights: Train the trainer programme, TeSS

Data Stewardship

- Competencies and skills
- Curriculum building
- Community building: Data Steward Interest Group
- Training, a few examples
- Tool: ELIXIR Data Stewardship Wizard (DSW)
- Standards: FAIR Training Working Group

Quick intro Celia

- ## Disclaimer:
  - I am not an expert in RDM and I am not a Data Steward
  - You are the experts in that area!

- ## My expertise is in Training:
  - Programme Manager DTL Learning/ELIXIR-NL Training
  - Co-lead ELIXIR Europe Training Platform
  - Member of Executive Board of GOBLET
  - My background: PhD in Biochemistry/Bioinformatics, Bioinformatics teacher, Training Manager, Coordinator, Organizer for many years

- ## I will share with you the ELIXIR projects that relate to data stewardship & training
  - You may notice some emphasis on the work of my own Node, ELIXIR-NL

ELIXIR: European infrastructure for biological information

Data infrastructure for Europe's life-science research:

22 Members and 1 Observers
~ 180 institutes involved
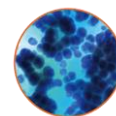
Data

Interoperability

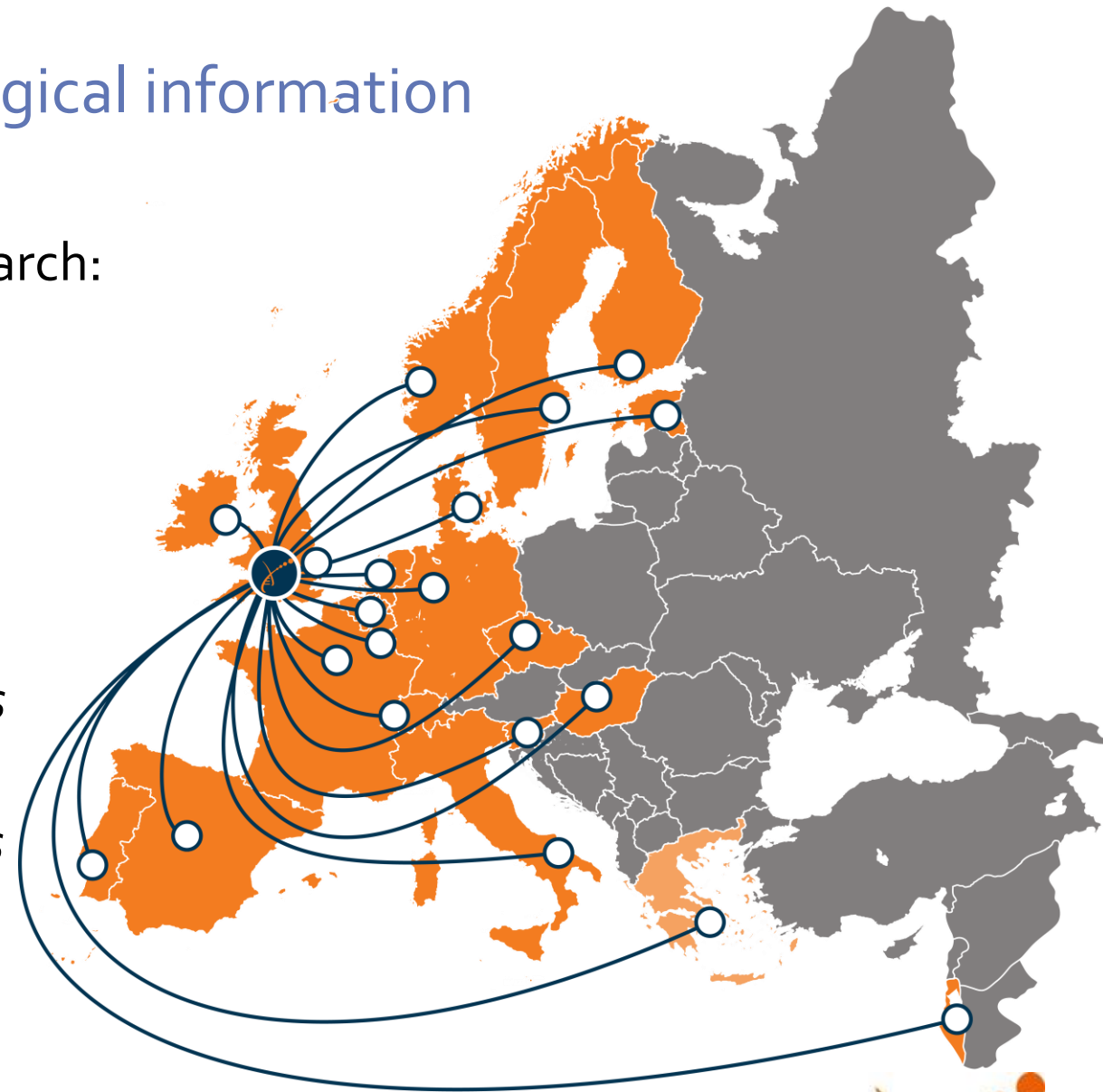Tools

Compute

Training

Marine metagenomics

*Crop and forest plants*

Human data

Rare diseases

www.elixir-europe.org

@ELIXIREurope

ELIXIR Training Platform

- **Build a sustainable training infrastructure** - supported and adopted by all member states

- **Deliver a coherent, high-quality and impactful ELIXIR training program**

  - addressing training gaps

  - growing bioinformatics capacity and competence in Europe

  - empowering researchers across Europe to use the ELIXIR services and tools

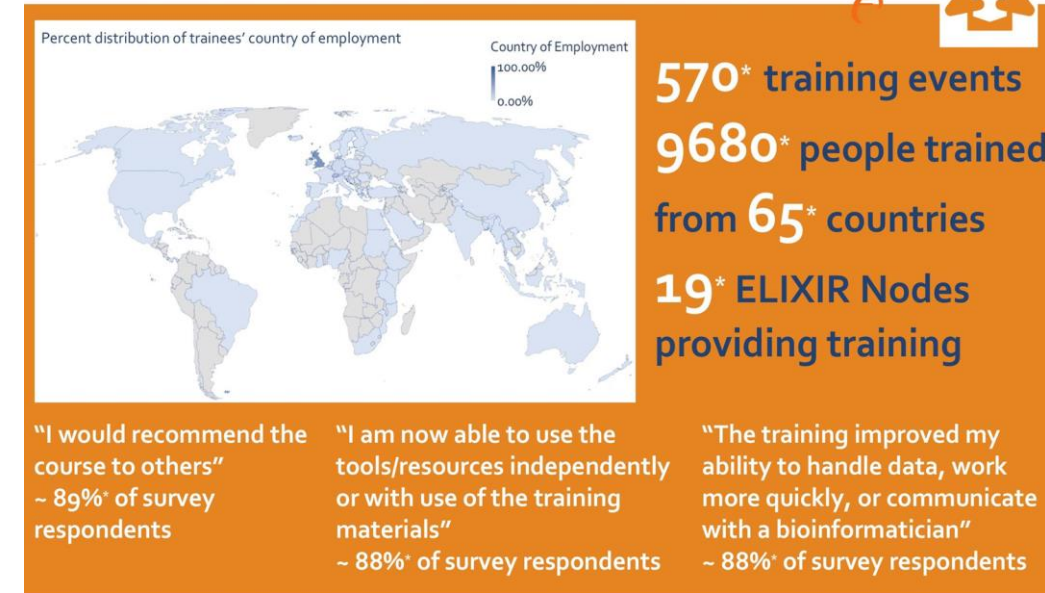- **Builds on and complements trainings in ELIXIR Nodes**
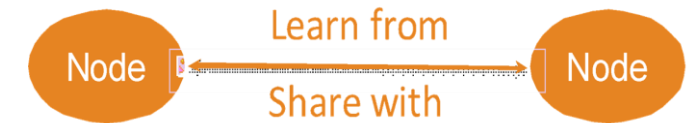
ELIXIR Training - in summary

- **Training events** – researchers, developers, infrastructure operators & trainers

- **Train-the-Trainer** - programme & developed framework

- **TeSS** portal - share & exchange training content and events *https://tess.elixir-europe.org*

- **Training Toolkit -** Training best practices, guidelines, standardized metrics for quality & impact assessment

- Solutions to ease training and communication - **E-learning platform** &**VCR**

Percent distribution of trainees' country of employment

Country of Employment
100.00%
0.00%

570* training events
9680* people trained
from 65* countries
19* ELIXIR Nodes providing training

"I would recommend the course to others"
~ 89%* of survey respondents

"I am now able to use the tools/resources independently or with use of the training materials"
~ 88%* of survey respondents

"The training improved my ability to handle data, work more quickly, or communicate with a bioinformatician"
~ 88%* of survey respondents

ELIXIR Training: Quality and Impact
https://www.elixir-europe.org/platforms/training

eXcelerate
*Sept 2015 - May 2018

Sept 2015 – May 2018

19 Nodes involved:
BE, CH, CZ, DE, EBI, EE, ES, FI, FR, GR, IL, IT, LU, NL, NO, PT, SE, SI, UK
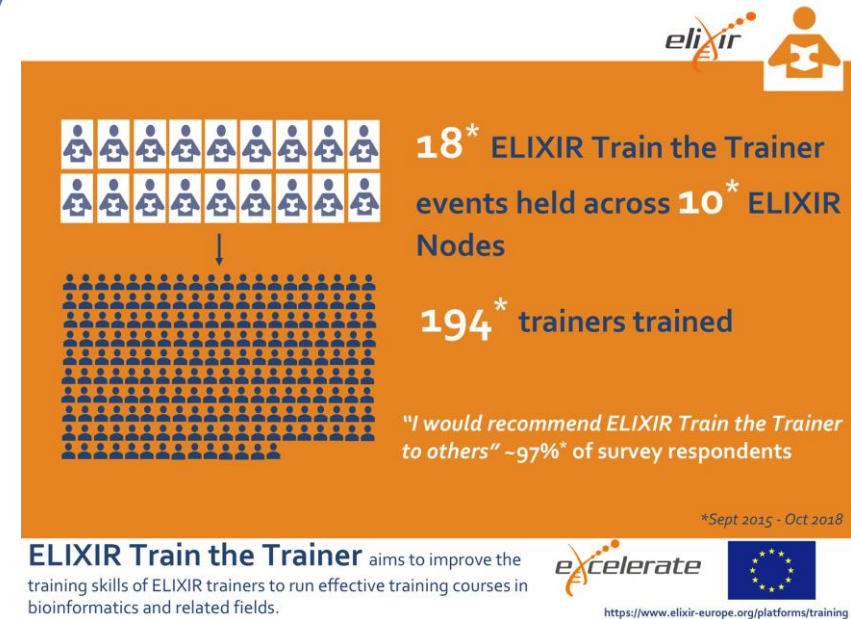
Train-the-Trainer (TtT) - Building a highly-skilled and coherent community of trainers

- Created an ELIXIR TtT programme with a set of materials and course layouts

- Created a TtT Exchange program – support trainers and Nodes

18* ELIXIR Train the Trainer events held across *10** ELIXIR Nodes

194* trainers trained

"I would recommend ELIXIR Train the Trainer to others" ~97%* of survey respondents

*Sept 2015 - Oct 2018

ELIXIR Train the Trainer aims to improve the training skills of ELIXIR trainers to run effective training courses in bioinformatics and related fields.

https://www.elixir-europe.org/platforms/training

Sept 2015 – May 2018

ELIXIR TtT Course setup

The ELIXIR TtT courses are organized around four sessions:

- Learning principles and how they apply to training
- Training techniques that can be used to enhance learner engagement and participation
- Assessment and feedback in training
- Session, course, and materials design

The Learning Objectives are:

- To get acquainted with learning principles and how they apply to training
- To be able to select and use training techniques that can enhance learner engagement and participation
- To learn how to use assessment and feedback in training
- To learn how to design a session, a course, and course materials

- https://github.com/TrainTheTrainer/EXCELERATE-TtT
- Contact: allegra.via@uniroma1.it & patricia.palagi@sib.swiss

ELIXIR Training Portal TeSS

TeSS is an aggregator

- 1242 training materials & 286 upcoming events

- 48 content providers (not only ELIXIR Nodes!)

- majority scraped automatically

- AAI and Bioschemas integration

- Linked to other ELIXIR registries: bio.tools and FAIRsharing

- Events annotated with EDAM, working on annotating with competencies to create "learning paths"

- TeSS License: BSD-3 for the Code and CC-BY 4.0 for the data

- Contact niall.beard@manchester.ac.uk for more information

https//tess.elixir-europe.org
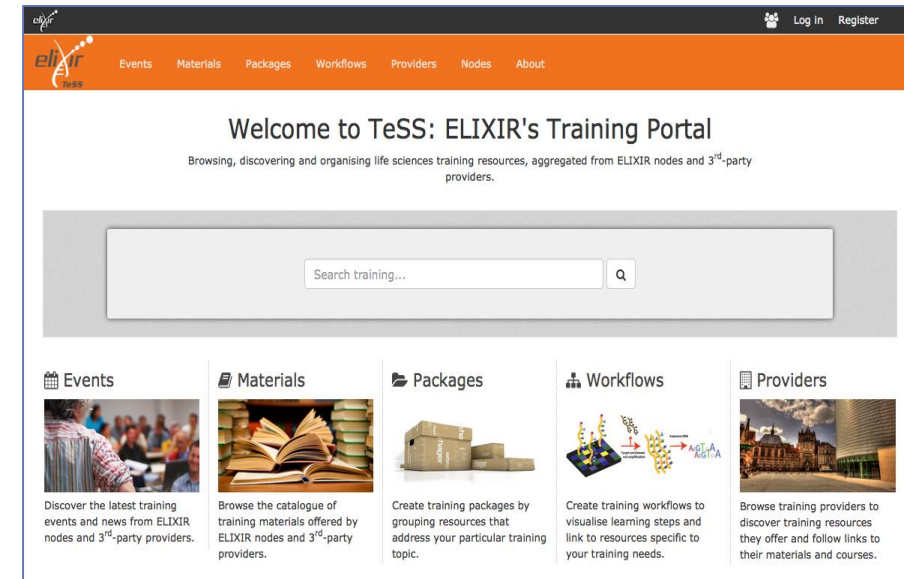
Data Stewardship related activities in ELIXIR
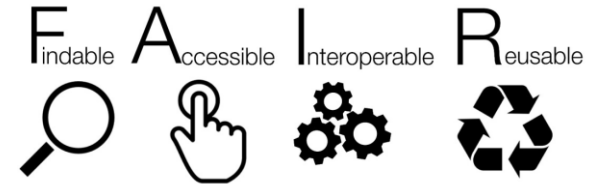
Findable Accessible Interoperable Reusable

- Competencies and skills
- Curriculum building
- Community building: Data Steward Interest Group
- Training, a few examples
- Tool: ELIXIR Data Stewardship Wizard (DSW)
- Standards: FAIR Training Working Group

- Most activities are across projects and across the ELIXIR Nodes
- Several activities connect to domains outside Life Sciences

Data Stewardship (DS)

- **Data Stewardship**
  Responsible planning and executing of all actions on digital data before, during and after a research project, with the aim of optimizing the usability, reusability and reproducibility of the resulting data
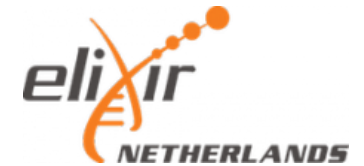
- **Capacity building in data stewardship is needed**
  - Skills in data management & data stewardship are crucial to deal properly with the data
  - Furthermore DMPs are obligatory for many if not all funding applications
  - DS has many many many aspects (nobody can do nor has to do all steps of DS)

Training in Data Stewardship – to whom (& what to teach precisely?)

- **Researchers**: create, manage, analyse and share FAIR data

- **Data Stewards**: provide support for research teams, data repositories, data services

- **Reviewers and funders**: assess DMPs, assess FAIRness of datasets, data resources in publications, grant proposals and reports

- **Trainers**: teach how to apply FAIR principles and how to perform good Data Stewardship

F_{indable} A_{ccessible} I_{nteroperable} R_{eusable}

elixir
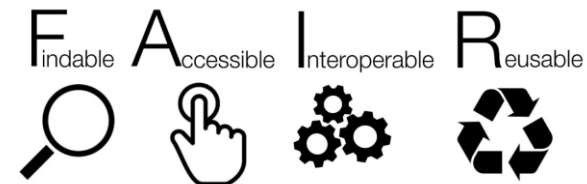
Competencies and skills:
Towards FAIR Data Steward as profession for the Life Sciences

- A collaborative approach built on existing expertise

- Aim: professionalise the data steward function within the life-sciences domain, with endorsement of all stakeholders in NL

- https://zenodo.org/communities/nl-ds-pd-ls/

- Contact: Salome Scholtens (UMCG), Celia van Gelder (DTL)

First deliverable – Life sciences data steward function matrix

- February 2019, *DOI:10.5281/zenodo.2561723*

- Created by
  a) analyzing ca 45 job advertisements and
  b) analyzing and mapping to existing data stewardship competency frameworks (EOSCpilot, Purdue, DAMA, EDISON)

- For 2 types of data stewards:

  - Data Steward A is an institutional/department focused and coordinating, policy-oriented function

  - Data steward B is project focused and an operational, supporting-oriented function

Life sciences data steward function matrix

Nine different function areas were defined for both data steward A & B:

- 1. policy/strategy,
- 2. compliance,
- 3. alignment with FAIR data principles,
- 4. services,
- 5. infrastructure,
- 6. knowledge,
- 7. network,
- 8. data archiving
- 9. supervision.

- Matrix well received, both nationally and internationally. It is a living document ! At the moment we are incorporating all feedback from the community

Ongoing Work

- Definition of the knowledge base, skills and
  learning objectives ,
  DOI: https://doi.org/10.5281/zenodo.2616477

- Next steps:
  - Education line,
    training material, pilot run,
    script for an eLearning module,
    sustainability plan to ensure
    national embedding

- **Save the date:**
  two 1-day Workshops 11 & 12 June 2019, Utrecht

Table 2: Example of learning objectives following Bloom's Taxonomy

RESPONSIBILITY:
Compliance

KNOWLEDGE:
Knowledge about legislation, ethics, code of conducts and research codes with regard to (research)data, (medical) research and privacy

SKILL/ABILITY:
Translate legislation and codes of conducts with regard to research data to practical implications and guidelines that researchers can understand

LEARNING OBJECTIVES:
-Explain the legislation and codes of conduct with regard to research data to researchers (applying)
-Examine and list the practical implications of legislation and codes of conduct with regard to research data (analyzing)
-Create guidelines based on legislations and codes of conduct with regard to research data (creating)

Figure 1: Venn diagram of different data steward types
Figure acknowledgement: Christine Staiger, DTL

16
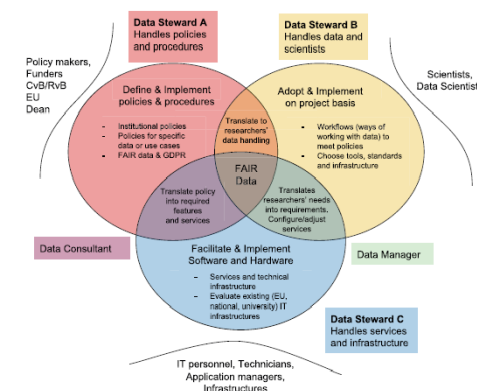
Community Building –
The Data Stewards Interest Group

Jasmin K. Böhmer — UMC Utrecht Center for Molecular Medicine
Marc Thompson — Leiden UMC Human Genetics
Christine Staiger — Dutch Tech Center for Life Science
Marije van der Geest — UMC Groningen Dept. of Genetics
Martiene Moester — Leiden UMC Biomedical Data Science
Miriam Braskova — Erasmus Research Institute of Management
Chrissie Taselaar — Princess Maxima Center for Pediatric Oncology

- June 2017 initial meeting of the Interest Group

- Leiden UMC, DTL, and UMC Utrecht founding members; mailing list has 90 members

  (NL, BE, LU, UK, SE, FR, DE)

- 9x meetings up until now;

  next meeting June 14, 2019 (remote attendance possible)

- Chair: Jasmin Böhmer, UMCU

- https://www.dtls.nl/community/interest-groups/data-stewards-interest-group/

Slide: Jasmin K. Böhmer

Mission & Activities of the Data Stewards Interest Group

Providing
... platform for practical experience exchange

Fostering
...the national implementation of Data Stewards(hip)

Joining
... efforts to produce hands-on solutions

The Input... experience, guidance, agreements, anecdotes, definitions

The Output ... use cases, documentation, case studies

So far:
- Collection of 24 stories and anecdotes for the promotion of good data stewardship
- Collaborative Paper on FAIR readiness for organizations
- Started inventory of training materials and courses

Slide: Jasmin K. Böhmer

Curriculum building & lesson development

- **Fair data stewardship curriculum :**
  - Lessons developed around the FAIRification process
  - Lesson objectives directly connected to the FAIR metrics
  - Modelled after the Carpentries,
  - will be linked to/from relevant parts of Data Stewardship Wizard

- **4OSS - Four simple recommendations to encourage best practices in research software**
  - *https://f1000research.com/articles/6-876/v1; https://github.com/SoftDev4Research/working-group*
  - Develop **open and reusable training materials** to teach researchers and developers how to adopt the 4OSS as development practices. First lesson release: *doi.org/10.5281/zenodo.2565040*
  - **Upcoming trainings: May 15, Ghent, Belgium and July 15-17, Athens, Greece**

- Contact: Mateusz.Kuzak@dtls.nl

Researchers are afraid of putting their code out there.

- What if I write crap code that nobody likes?

- Will people judge me for the code I write?

- What if someone finds a bug in my code?

- What if I get scooped?

- What are the challenges of Open Software Development?

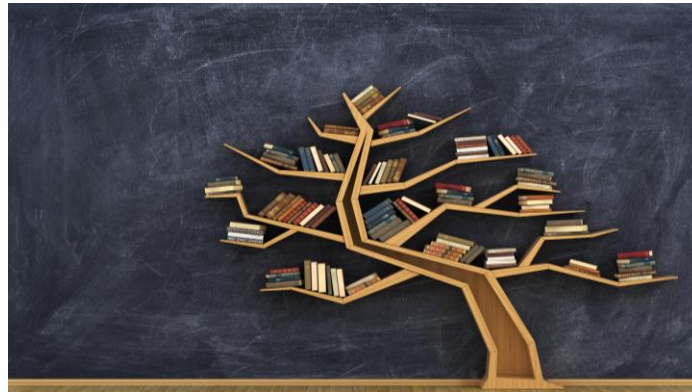- What if someone wants to contribute to my code?

1 OPEN SOURCE YOUR CODE FROM DAY ONE
Make your source code publicly accessible in a version-controlled repository (e.g. github.com and bitbucket.org) and increase reproducibility, reusability and collaboration.

2 MAKE YOUR SOFTWARE DISCOVERABLE
Register your software metadata in a popular community registry (e.g. bio.tools) and increase your project's visibility.

3 MIND THE LICENSE
Adopt a license that specifies how others can use and distribute your software. Ensure that the software fits with the license of third-party dependencies.

4 DEFINE RESPONSIBILITIES
Let people know how they can contribute to your project and contact you.

Training example 1:
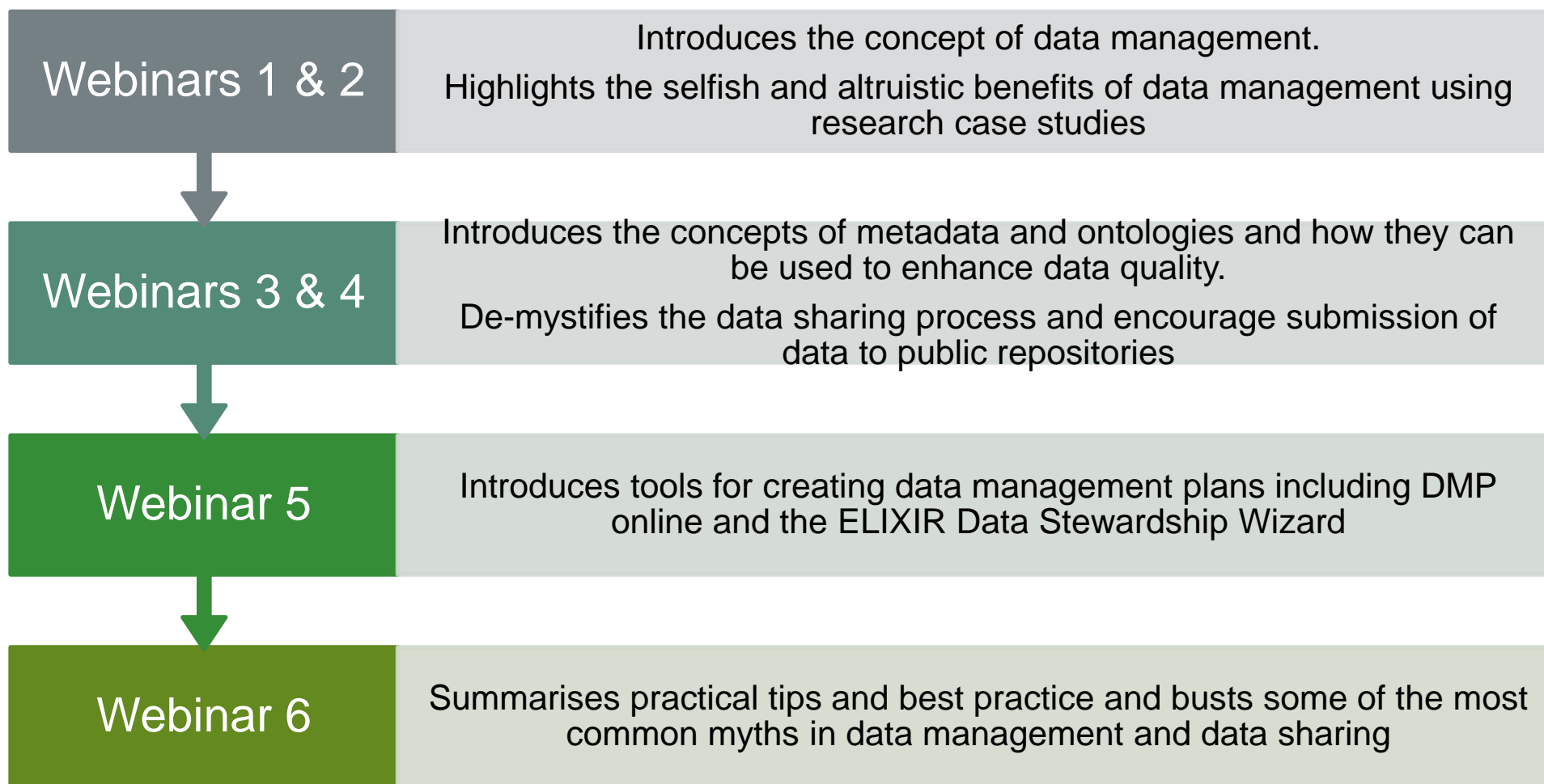# Bringing data to life: data management for the biomolecular sciences

- ELIXIR-EBI webinar series
- 28 June – 2 August 2018; Watch the series in Train online https://bit.ly/2NQ9tXS
- Audience: Biomolecular researchers
- Content: Six webinar featuring experts from EMBL-EBI, TU-DELFT, Wellcome Sanger Institute, DCC and ELIXIR
- Contact: Melissa Burke (mburke@ebi.ac.uk) and Sarah Morgan (sarahm@ebi.ac.uk)

Content in the webinars

Webinars 1 & 2	Introduces the concept of data management. Highlights the selfish and altruistic benefits of data management using research case studies
Webinars 3 & 4	Introduces the concepts of metadata and ontologies and how they can be used to enhance data quality. De-mystifies the data sharing process and encourage submission of data to public repositories
Webinar 5	Introduces tools for creating data management plans including DMP online and the ELIXIR Data Stewardship Wizard
Webinar 6	Summarises practical tips and best practice and busts some of the most common myths in data management and data sharing

Training Example 2: Carpentries workshops

- Pilot Software and Data Carpentry in 2015/16: 278 life sciences researchers and 63 instructors trained

- Since 2017:

  - \> 40 ELIXIR courses have been delivered, many more delivered by the ELIXIR Nodes themselves

  - ELIXIR/Carpentries agreement to support the rolling out across Nodes:

    - 20 Software Carpentry and Data Carpentry workshops

    - 2 Instructor trainings

  - Nodes are building up their Carpentries trainer pools

- Contact allegra.via@uniroma1.it and pkahlem@googlemail.com

✉ Aleksandra Pawlik[1], ✉ Celia W.G. van Gelder (iD)[2], ✉ Aleksandra Nenadic (iD)[3], Patricia M. Palagi[4], Eija Korpelainen[5], Philip Lijnzaad[6], Diana Marek[4], Susanna-Assunta Sansone[7], John Hancock[8], Carole Goble[3]

Training Example 3: Helis Academy

- Postgraduate courses for Data Analysis and Stewardship for Industry and Academia
- Bridge gap between industry needs and existing training programs and curricula
- Training program that
  - Improves industry-readiness of graduate students in the Health and Life Sciences
  - Upskill employees (life-long learning) in the Health and Life Sciences sector

Good Manufacturing Practices (GMP)
Clinical Testing
Data Analysis and Stewardship
Business Development
Technology & Life Sciences
Product and Process Design

Partners

- *https://www.helisacademy.com/en*
- Contact: Christine Staiger & Celia van Gelder (DTL/ELIXIR-NL), Friederike Ehrhart, Lauren Dupuis (Maastricht University)

Helis Academy

Helis Academy Data Analysis and Data Stewardship Learning Blocks

Omics Data Analysis Transcriptomics data analysis Multi omics integration	**FAIR Data Stewardship** Data management planning Research life cycle
Statistics Statistical thinking Experimental design	**Machine Learning** Visual analytics for genomics Data mining

- Course FAIR Data Stewardship: 27-29 May 2019
  https://www.aanmelder.nl/helis-fair-data-stewardship-course
- Course Omics Data Analysis: 11-14 June 2019
- Location: Niel, Belgium
- Courses can be followed together or separately
- Courses will be run in Belgium and Netherlands in year 2 and 3 of the project

Tool: The ELIXIR Data Stewardship Wizard

- The Data Stewardship Wizard (DSW) is a web-based application for preparing smart Data Management Plans.

- It connects to additional learning resources and helps raise awareness about DS challenges.

- *https://ds-wizard.org/*

The Wizard is for:

- **Researchers:** Support in DS Planning, Raise awareness

- **Data Stewards:** Check list DS Planning

- **Experts:** Their contact details can be linked

- **Funders / Institutes:** Customize, Domain specific

- Allows tailor made DSW instances for an organisation, a Node, a company etc.

- Contact: rob.hooft@dtls

Standards: ELIXIR FAIR Training Working Group

- chair: Mateusz Kuzak (ELIXIR-NL)

- Task force 1: **Training about FAIR data**
    - making training resources findable via TeSS, the ELIXIR Training Portal
    - Defining necessary skills for implementing FAIR data
    - developing a set of structured terms to classify the skills necessary to make data FAIR and keep it FAIR (with GO-TRAIN)
    - lead: Mateusz Kuzak (ELIXIR-NL), Fotis Psomopoulos (ELIXIR-GR)
- Task force 2: **Making training resources FAIR**
    - developing and implementing FAIR metrics for training resources ((in collaboration with Michel Dumontier and the FAIR metrics group)
    - lead: Leyla Garcia (ELIXIR-EBI), Bérénice Batut (ELIXIR-DE)

elixir

Summary

Highlights of ELIXIR Data Stewardship related activities shown

- Competencies and skills
- Curriculum building
- Community building: Data Steward Interest Group
- Training, a few examples
- Tool: ELIXIR Data Stewardship Wizard (DSW)
- Standards: FAIR Training Working Group

- We welcome feedback and also invite you to participate in our activities!
- All is open for use and reuse

Acknowledgements

Thank you!