

## Deliverable D5.3 Report on the dedicated training and capacity building activities

<b>Project Title (Grant agreement no.):</b>	ELIXIR-CONVERGE: Connect and align ELIXIR Nodes to deliver sustainable FAIR life-science data management services (871075)		
<b>Project Acronym (EC Call):</b>	ELIXIR-CONVERGE (H2020-INFRADEV-2018-2020)		
<b>WP No &amp; Title:</b>	WP5 Demonstrator Projects		
<b>WP leader(s):</b>	Anne-Françoise Adam-Blondon		
<b>Deliverable Lead Beneficiary:</b>	27 - UL		
<b>Contractual delivery date:</b>	28/02/2023	<b>Actual delivery date:</b>	24/04/2023
<b>Delayed:</b>	Yes		
<b>Partner(s) contributing to this deliverable:</b>	BSC, SIB, HITS, INRAE, CNRS, MTA, CNR, UNILU, UIB, INESC-ID, IGC, UU, UL, CRG		
<b>Authors:</b> Brane Leskošek (UL), Vanita Haurheeram (INRAE), Federico Bianchini (UiB/UiO), Pinar Alper (UNILU), Laura Portell-Silva (BSC), Espen Åberg (UiB/UiT), Adam Hospital (IRB-Barcelona), Marina Popleteeva (UNILU), Nils P Willassen (UiB/UiT). Marko Vidak (UL), Anne-Françoise Adam-Blondon (INRAE)			
<b>Contributors:</b> Olivier Collin (CNRS), Amy Curwin (BSC), Mallory Freeberg (BSC), Erwan Le Floch (INRAE), Paulette Lieby (CNRS), Ernesto Picardi (CNR), Salvador Capella-Gutierrez (BSC)			
<b>Acknowledgments (not grant participants):</b> Cyril Pommier (EMPHASIS, ELIXIR Plant Science community, INRAE, FR); Sebastian Beier (ELIXIR Plant science community, IPK, DE); Jérôme Grimplet (CA 17111 Integrape, CITA, SP); Camille Rustenholz (CA 17111 Integrape, University of Strasbourg, FR), Marco Brandizi (ELIXIR Plant science community, Rothamsted University, UK), Inês Chaves (ELIXIR Plant science community, ITQB, PT), Daniel Faria (ELIXIR Plant science community, University of Liboa, PT)			
<b>Reviewers:</b>	ELIXIR-CONVERGE Management Board (MB) members.		

### Log of changes

DATE	Mvm	Who	Description
15/02/2023	0v1	Brane Leskošek (UL)	Initial version
07/04/2023	0v2	A-F Adam-Blondon (INRAE)	Sent to PMU after incorporating internal WP feedback
14/04/2023	0v3	Nikki Coutts	Circulated to the MB for final review before

		(ELIXIR Hub)	submission
24/4/2023	1v0	Nikki Coutts (ELIXIR Hub)	Final version to be uploaded into EC Portal

## Table of contents

<b>1. Executive Summary</b>	<b>2</b>
<b>2. Contribution toward project objectives</b>	<b>2</b>
<b>3. Introduction</b>	<b>4</b>
<b>4. Description of work accomplished</b>	<b>4</b>
4.1. Contribution of WP5 to WP2's inventory of training material	4
4.2 Collection of training needs through 1-2-1 interviews	5
4.3 Joint organisation of a workshop by WP2, WP5 and the ELIXIR training platform to complete and structure training needs	5
4.4 Participation to the 2022 ELIXIR Biohackathon to develop a learning path	5
4.4.1 Development of a first version during the 2022 ELIXIR Biohackathon	5
4.4.2 Development of an improved version with members of the ELIXIR Plant Science community	6
4.4.3 Mapping existing training material on the learning path and assessing its FAIRness	6
<b>5. Results</b>	<b>7</b>
5.1. Contribution of WP5 to WP2's inventory of training material	7
5.2 Identification of training gaps in the context of the use cases	8
5.2.1 Surveys and experts interviews	8
5.2.2 Workshop to complete and structure training needs	9
5.3 Development of learning paths	10
5.3.1 Federated access to human genomics data: GDPR	10
5.3.2 Harmonised FAIR plant genotype & phenotype data management toolkit for Europe	11
<b>6. Conclusions</b>	<b>16</b>
<b>7. Impact</b>	<b>17</b>
<b>8. Next Steps</b>	<b>17</b>
<b>9. Deviation from Description of Action</b>	<b>17</b>

## 1. Executive Summary

One of the objectives of WP5 was to identify the needs in terms of capacity building, in the context of



the 6 use cases, and to develop strategy(ies) in collaboration with WP2 to address these needs. This was achieved in three steps:

- Create an inventory of training resources available in the context of the use cases
- Send an online survey followed by a series of F2F meetings with each use case to collect their training needs
- Test the development of learning paths as a method to develop a structured overview of training needs in relation to a use case

The method set up by the ELIXIR Training Platform to better structure the development of training resources was proved very helpful to specify the needs in terms of training and capacity building in the context of the use cases. Spelling out the priority in terms of training target and the learning path to follow to reach this target is a way to detail the different modules that are necessary. It is then possible to map on these modules the training resources that are already available and to derive a list of gaps. This method was applied for three demo cases successfully in the Plant, Marine metagenomics and Human Genome data contexts.

In terms of general organisation of community resources to facilitate this type of approach, the most important actions are to make the training materials accessible for reuse and easy to find. Versioning is highly desirable, to facilitate collaborative development and maintenance of training material of the time and a DOI attribution will facilitate citation and long term findability. Findability requires additional work to organise the training events and training resources in categories and/or facilitate filtering in the context of catalogues of training resources such as [TeSS](https://tess.elixir-europe.org/)<sup>1</sup> or [Glitr](http://glitr.org)<sup>2</sup> or of knowledge hubs such as the [RDMkit](https://rdmkit.elixir-europe.org/)<sup>3</sup>. This is achieved in TeSS, through the annotation of the training material and linked training events with ontology terms from EDAM, as well as with custom keywords.

## 2. Contribution toward project objectives

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

Objective no. / Key Result no. Description	Contributed to:
<b>Objective 1:</b> Develop a sustainable and scalable operating model for transnational life-science data management support by leveraging national capabilities ( <b>WP1, WP5</b> )	
<b>Key Result 1.1:</b> Established European expert network of data stewards that connect national data centres and similar infrastructures and drive the development of interoperable solutions following international best practice, including national interpretations of the General Data Protection Regulation (GDPR)	Yes
<b>Key Result 1.2:</b> Development of joint guidelines and common toolkit that are adopted into funder recommendations, with support available nationally and in	Yes

<sup>1</sup><https://tess.elixir-europe.org/>

<sup>2</sup><http://glitr.org>

<sup>3</sup><https://rdmkit.elixir-europe.org/>

local languages	
<b>Key Result 1.3:</b> The catalogue of successful national business models incorporated into national strategies	<b>No</b>
<b>Key Result 1.4:</b> The developed “sustainable and scalable operating model for transnational life-science data management support” is adopted into national ELIXIR Node	<b>No</b>
<b>Objective 2:</b> Strengthen Europe’s data management capacity through a comprehensive training programme delivered throughout the European Research Area ( <b>WP2, WP5</b> )	
<b>Key Result 2.1:</b> A comprehensive ELIXIR Training and Capacity building programme in Data Management, directed at both data managers and ELIXIR users, and connected to the national training programmes in Data Management in the ELIXIR Nodes and prospective ELIXIR Member countries.	<b>Yes</b>
<b>Key Result 2.2:</b> Development of a collective group of trainers that support scalable deployment of Data Management training across ELIXIR Nodes.	<b>Yes</b>
<b>Key Result 2.3:</b> A substantial cohort of data managers, Node coordinators and researchers with specific data management skills, business planning and knowledge of transnational operations across the ELIXIR Nodes	<b>No</b>
<b>Objective 3:</b> Align national data management standards and services through a sustainable, scalable and cost-effective data management toolkit ( <b>WP2, WP3, WP5</b> )	
<b>Key Result 3.1:</b> Assemble a full-stack harmonised common toolkit comprising all aspects of data management: from data capture, annotation, and sharing; to integration with analysis platforms and making the data publicly available according to international standards.	<b>No</b>
<b>Key Result 3.2:</b> Provide exemplar toolkit configurations for prioritised demonstrators to serve as templates for future use.	<b>No</b>
<b>Key Result 3.3:</b> Establish national capacity in using as well as updating, extending and sustaining the toolkit across the ERA.	<b>No</b>
<b>Key Result 3.4:</b> Enable ‘FAIR at source’ practice for data generation, and analytical process pipeline implementation by flexible deployment of the toolkit in national operations	<b>No</b>
<b>Objective 4:</b> Align national investments to drive local impact and global influence of ELIXIR ( <b>WP4,WP6</b> )	
<b>Key Result 4.1:</b> Development of a Node Impact Assessment Toolkit based on RI-PATHS methodology.	<b>No</b>
<b>Key Result 4.2:</b> Adoption of Impact assessment in ELIXIR Nodes, supported by	<b>No</b>

Node coordinators network and feedback on applicability from dialogues with national funders.	
<b>Key Result 4.3:</b> Creation of national public-private partnerships and industry outreach where open life-science data and services stimulate local bioeconomy	<b>No</b>
<b>Key Result 4.4:</b> Growth in reach, impact and engagement of stakeholder communication assessed by established ELIXIR Communications metrics	<b>No</b>
<b>Key Result 4.5:</b> Initiating and advancing discussions on Membership (EU and international) or strategic partnerships (international countries) following ELIXIR-CONVERGE workshops.	<b>No</b>

### 3. Introduction

The goal of ELIXIR-CONVERGE WP5 is to increase the capacity of ELIXIR and its national nodes to implement Data Management Plans (DMPs) in their projects at an EU scale. To assist the development of resources and support this objective, six demonstrator use-cases were chosen at the start of the project:

1. Harmonised FAIR plant genotype & phenotype data management toolkit for Europe
2. Reproducible, comparable and FAIR Epitranscriptomics
3. Common Data Management Plans for the Marine Metagenomics Community
4. Federated access to human genomics data: GDPR
5. FAIR encoding and access to Toxicology data
6. FAIR organisation of biomolecular simulation information

One of the objectives of WP5 was to identify the needs in terms of capacity building, in the context of the 6 use cases, and to develop strategy(ies) in collaboration with WP2 to address these needs. This was achieved in three steps:

- Create an inventory of training resources available in the context of the use cases
- Send an online survey followed by a series of F2F meetings with each use case to collect their training needs
- Test the development of learning paths as a method to develop a structured overview of training needs in relation to a use case

## 4. Description of work accomplished

### 4.1. Contribution of WP5 to WP2's inventory of training material



This activity completed WP2's inventory done at the start of the project with material specific to the use cases domains. In the context of WP5, the survey was started at the beginning of the project in parallel to WP2 survey and completed during a virtual workshop on training gaps organised jointly by WP2, WP5, and the ELIXIR Training platform.

Representatives from WP5 also attended the WP2-WP3 hackathon (two sessions: 23<sup>rd</sup> September and 18<sup>th</sup> October 2021) to register and annotate courses on TeSS and to enable linkage between TeSS and RDMkit.

## 4.2 Collection of training needs through 1-to-1 interviews

The training needs for the different demonstrator use-cases were identified through a series of 1-to-1 interviews with experts involved in the demonstrator use-cases. To get context and prepare the interviews, a survey prepared in collaboration with WP2 was sent previously to the representatives of the demonstrator use-cases (Annexe 1). This survey wanted to capture an initial idea of already available and missing training courses and materials for each demonstrator use-case.

## 4.3 Joint organisation of a workshop by WP2, WP5 and the ELIXIR training platform to complete and structure training needs

A virtual workshop on training gaps was organised in two iterations (June 15<sup>th</sup> and June 29<sup>th</sup>, 2022). The objectives of these events were the following:

- Learn about collections of FAIR training material and Learning Paths developed by the ELIXIR Training platform
- Continue to identify already available training materials for the use-cases domains.
- Continue to identify training gaps in the use-cases domains.
- Identify training priorities and learning objectives for future courses and training materials.

During this workshop, five out of the six use-cases started to structure their needs in terms of training.

## 4.4 Participation to the 2022 ELIXIR Biohackathon to develop a learning path

### 4.4.1 Development of a first version during the 2022 ELIXIR Biohackathon

The ELIXIR training platform developed resources to facilitate the development of Learning Paths and proposed to test them during a track of the 2022 ELIXIR BioHackathon: presentation of the concepts of FAIR collections of training resources and of learning paths; short e-learning videos on specific aspects of the learning paths and a template for developing new learning paths. This opportunity was used to develop a learning path in the context of the Demonstrator Project "Harmonised FAIR plant genotype & phenotype data management toolkit for Europe".

The key steps of the development of a learning path are the following:

- Identify the domain and the target audience.
- Identify areas within this domain which need to be covered during the learning path.
- Identify topics that can be mapped on the targeted areas and define the required stages of expertise of the learners (from beginner to advanced) and the cognitive level to be reached after training according to Bloom's taxonomy (the Bloom's cognitive taxonomy consists of six conceptually different levels within the cognitive domain : Remember, Understand, Apply, Analyse, Evaluate and Create).
- Connect the topics based on the prerequisites with an arrow, the topic from which the arrow originates is a prerequisite for the topic towards which this same arrow points.
- Identify the start points (without incoming arrow) and the end points (without outgoing arrow) of the learning path.
- Annotate the learning outcomes of the learning path that represents the overall outcome of each end point.

A first version of the learning path was developed by two ELIXIR-CONVERGE partners following this methodology and with the support of the training platform members.

#### 4.4.2 Development of an improved version with members of the ELIXIR Plant Science community

A set of other members of the Plant Science community helped to refine the first draft based on their own training experience and needs.

#### 4.4.3 Mapping existing training material on the learning path and assessing its FAIRness

The next step was to gather training materials that can be mapped onto the different topics identified. The training materials found came from different sources, some from a file that gathers the training materials of each partner of this WP (see section 5.1), others from TeSS<sup>4</sup> and some others from collaborators of the plant community that were surveyed during the workshop organised by WP5.

The following stage consisted in rating them in terms of FAIRness, as described in the article "Ten simple rules for making training materials FAIR" by Garcia L, Batut B, Burke ML, Kuzak M, Psomopoulos F, et al (2020) <https://doi.org/10.1371/journal.pcbi.1007854>. Selected criteria to assess this rating are presented in Table 1.

**Table 1.** Suggested metadata for FAIR training events and associated materials. Metadata related to the training material itself are proposed in accordance with the practices recommended by the ELIXIR Training platform and implemented on the TeSS training portal. Metadata related to the training event are taken from the article cited above.

FAIR criteria	Description of the training event and material metadata
Title	Title of the event and associated training material.
Contact details	Author(s) name of the training material and names of the coordinators of the event

<sup>4</sup><https://tess.elixir-europe.org/>

FAIR criteria	Description of the training event and material metadata
	together with their contact details.
Description	Overview of the subject matter, aims of the training, and language in which the training is delivered.
Keyword	Keywords or tags identifying the topic of the training event and associated materials.
Target audience	The intended audience, their prerequisite knowledge and skills, their general background, and how the training material will help them.
Learning outcomes	Statements that indicate what trainees should be able to do upon successful completion of the training.
Prerequisites	Skills required before following training.
Persistent identifier	Provide a persistent identifier to the training material (e.g persistent uniform resource locator; digital object identifier)
Availability in a repository (share)	Training material shared in a suitable data repository, preferably in cross-domain trusted repositories (e.g., Zenodo, GitHub)
Availability in TeSS (findability)	Training material registered and annotated in TeSS (ELIXIR's training portal)
Licensing and (re)use details	Licence under which the materials are shared, and rules and conditions for (re)use and contribution.
Preferred citation	Instructions on how to cite your material.
Required resources	Technical resources and related materials (software requirements, datasets, infrastructure requirements, etc.).
Structure and duration	Description of the structure of the materials and setting in which to deliver them, including the time allocated to each part (lectures, exercises, etc.).
Additional information	Items that provide additional information about (re)use and delivery of the materials (e.g., general tips and guidance).
Links and references	Links and references that are relevant to the content but not required for delivery of the materials.
Date of last revision	Date of last update of the materials and the version.



## 5. Results

### 5.1. Contribution of WP5 to WP2's inventory of training material

17 training events related to the CONVERGE WP5 demonstrator use-cases (organised by 10 Nodes with more than 150 participants) are listed in the ELIXIR-CONVERGE WP5 Training Inventory (Annexe 2)<sup>5</sup>. Among them, 14 have open-access training materials.

Some WP5 partners also contributed to register events and material on TeSS (if not registered already) and to improve their annotation in a hackathon organised jointly by WP2 and WP3. When possible, the EDAM ontology was used for the categories “Scientific Topics”, where the domains can be added, and “Operations”. When EDAM terms were not available, a set of keywords based on the existing pages of RDMkit was introduced to ensure a consistent notation between the events organised in the context of ELIXIR-CONVERGE. This activity resulted also in a gap analysis with EDAM initiated a closer collaboration between ELIXIR-CONVERGE and EDAM personnel to better describe concepts and operations related to research data management. Finally during this hackathon, the group worked on creating and testing RDMkit queries to automatically discover relevant training materials on TeSS directly from the RDMkit pages. In the current implementation of RDMKit, these queries are linked at the bottom of the pages in the “training” window.

### 5.2 Identification of training gaps in the context of the use cases

#### 5.2.1 Surveys and experts interviews

The survey and the 1-to-1 interviews results provided better insights into training needs, and will help develop and deliver target training events for selected demonstrator projects. The results of the 1-to-1 interviews with use-case representatives are summarised in table 2.

**Table 2.** Summary of the results of the survey and 1-2-1 talks on training and capacity building needs. **DP1:** Harmonised FAIR plant genotype & phenotype data management toolkit for Europe. **DP2:** Reproducible, comparable and FAIR Epitranscriptomics. **DP3:** Common Data management plans for the Marine Metagenomics Community. **DP4:** Federated access to human genomics data: GDPR. **DP5:** FAIR encoding and access to Toxicology data. **DP6:** FAIR organisation of biomolecular simulation information.

Demonstrator Project (DP)	DP1	DP2	DP3	DP4	DP5	DP6
Number of completed surveys	3	1	1	2	1	1
1-2-1 talks	Yes	Yes	Yes	Yes	Yes	Yes

<sup>5</sup> [https://docs.google.com/spreadsheets/d/1az4U22K4pPOnzbn\\_JcQy4bSMRaBKMGA0eHM2t07ZHeU/edit#gid=0](https://docs.google.com/spreadsheets/d/1az4U22K4pPOnzbn_JcQy4bSMRaBKMGA0eHM2t07ZHeU/edit#gid=0)

1-2-1 talks participants	Adam-Blondon, AF; Le Floch, E	Picardi, E	Aberg, E; Willassen, NP	Popleteeva, M	D'Altri, T.	Hospital, A
Existing training events	Limited	No	Limited	Limited	Sufficient	No
Existing materials	Limited	Limited	Limited	Limited	Sufficient	No
Potential trainers	Yes	No	Yes	Yes	Yes	Yes
Challenges	Build a DMP / knowledge model, plant genotyping	Computational workflows				FAIR simulation data, sharing of simulation data

### 5.2.2 Workshop to complete and structure training needs

#### Harmonised FAIR plant genotype & phenotype data management toolkit for Europe

Experts from the ELIXIR Plant Science community and from the [COST action CA17111](#)<sup>6</sup> were invited. During the workshop, the landscape of training resources was completed. As a first approximation to develop a learning path for the plant science community, the training resources were mapped to the [Plant Phenomic Tool Assembly](#)<sup>7</sup> elements to understand if there was any particular need of training in any of the parts. During the workshop, two learning objectives were identified:

1. How to make Plant phenotyping and genomic data interoperable for integration by machines, publication and reuse.
2. How to annotate and format Plant phenotyping and genomic data for publication of machine actionable knowledge.

The second objective is beyond the use case developed in ELIXIR-CONVERGE but will have to be addressed in the future.

#### Reproducible, comparable and FAIR Epitranscriptomics:

In the context of this use-case, the training needs are tightly linked to the evolution of the technologies used to produce the data and of workflows needed to analyse these data. The main learning objectives identified were:

1. Know methods to identify RNA modifications (type and precise position)
2. Know bioinformatics tools to detect RNA modifications.

<sup>6</sup><https://integrape.eu/>

<sup>7</sup>[https://rdmkit.elixir-europe.org/plant\\_phenomics\\_assembly](https://rdmkit.elixir-europe.org/plant_phenomics_assembly)

The main resource currently available to address these points is a publication in Nature Protocol (Lo Guidice et al 2020, <https://doi.org/0.1038/s41596-019-0279-7>) that could be transformed into a learning path.

#### Common Data management plans for the marine metagenomics Community:

The training needs identified by the marine metagenomics community members during the workshop are related to how to use the NeLS tool assembly. In addition, during the workshop other important needs for the community were identified:

1. A centralised repository for training material (e.g. on GitHub) linked to TeSS
2. Ways to more easily find relevant materials to assemble them in comprehensive courses and ways to store and find these courses.

#### Federated access to human genomics data: GDPR

During the workshop, the inventory of training resources was carefully reviewed and completed. A learning path was designed, with three domains and detailed learning outcomes (see details in section 5.1.1). The areas lacking training material were identified: gaps are mostly associated with advanced skills in the identification and management of ethical (GDPR) issues (see section 5.1.1).

#### FAIR organisation of biomolecular simulation information

The partners of the [BioExcel CoE](#)<sup>8</sup> have developed a lot of training resources, often associated to the documentation of the tools they develop and use. More recently, BioExcel has developed a [BioExcel Knowledge Resource Center](#)<sup>9</sup> aiming at gathering and starting to structure these resources. Unfortunately the training experts of the projects could not be involved in the discussion due to a funding gap at the moment of CONVERGE activities.

## 5.3 Development of learning paths

### 5.3.1 Federated access to human genomics data: GDPR

#### *Priority in terms of target audience.*

The target audience are data stewards and researchers developing research in the context of human biology or genomics and need to be compliant with GDPR.

#### *Learning Outcomes.*

---

<sup>8</sup><https://bioexcel.eu/>

<sup>9</sup><https://krc.bioexcel.eu/training>

The learning path is conceived with three domains that must be followed the specific order: the first one is a prerequisite for the second, which is a prerequisite for the third one:

- an introduction to data protection;
- a course on Records of Processing Activities in the context of research (ROPA);
- a course on Data Protection Impact Assessments in the context for research (DPIA).

For each domain/course, the learning outcomes are detailed below. The areas where few or no training materials exist were identified (tagged “GAP” below).

#### Data Protection in Research, an Introduction

- Learners can recognise and recall relevant GDPR vocabulary.
- Learners can list the GDPR requirements for accountable use of human data in research.
- Learners can describe the situations where the GDPR applies and does not apply.
- Given a research study, the learners can identify the relevant GDPR roles of participants of the study and know their responsibilities.
- Learners can list examples of possible technical and organisational data protection measures (GAP).
- Learners can tell the characteristics of data that determine whether data is personal data or not and the sensitivity.
- Learners can tell the characteristics of data use that constitute high risk data processing and a Data Protection Impact Assessment (DPIA) is needed.

#### Records of Processing Activities ROPA for research (prerequisite: Introduction)

- Learners can list the documentation they need to keep about research studies with human-subject data under the GDPR.
- Learners can list the audit trail they need to keep for research studies with human-subject data under the GDPR (GAP).
- Learners should list possible tools and the pros and cons of tools that can be used to implement a ROPA (GAP).
- Learners should know the ELIXIR resources that provide guidelines on data protection for research (GAP).
- Learners should possess practical (hands-on) experience with at least 2 ROPA tools from RDMkit (GAP).

#### Data Protection Impact Assessments DPIAs for research (prerequisite: Introduction and ROPA)

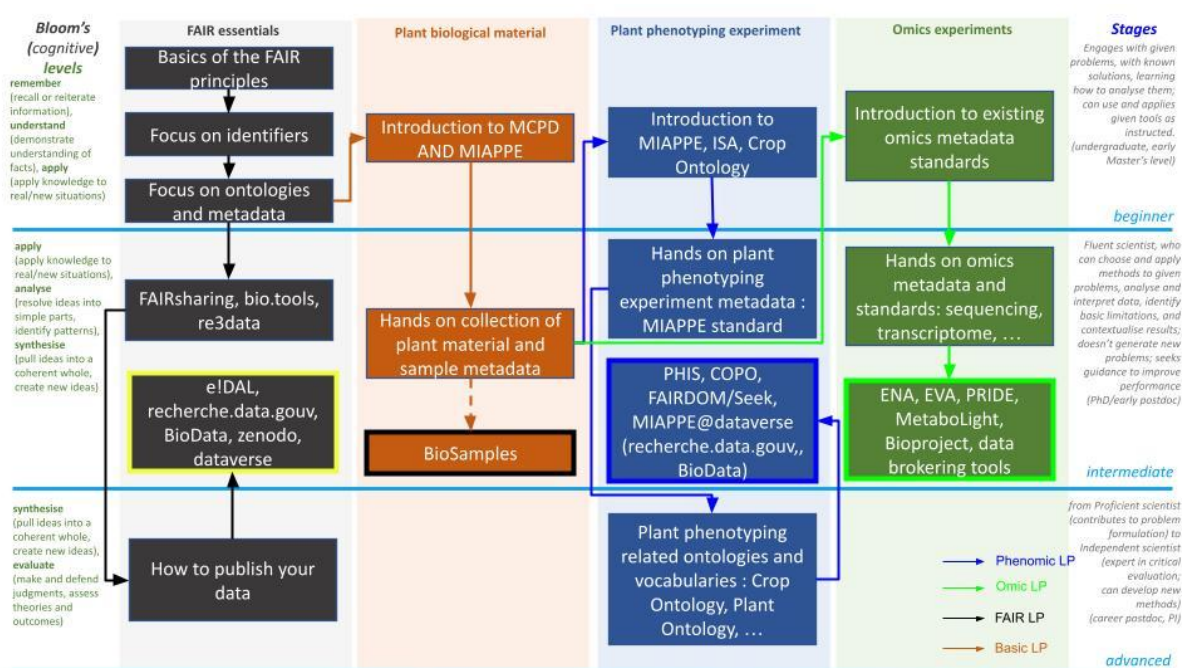
- Learners should list the characteristics of research studies that constitute a high risk of personal data process.
- Learners should know about different DPIA templates from data protection agencies.
- Learners should recall the basic aspects of risk assessments (GAP).
- Learners should list possible tools that can be used to perform a DPIA (GAP).

- Learners should know the difference between process and information risks and the different kinds of data protection measures (organisational and technical) to mitigate risks (GAP).
- Learners should possess (hands-on experience) with the DSW and the CONVERGE DPIA Knowledge Model to generate a DPIA for an example project (GAP).

### 5.3.2 Harmonised FAIR plant genotype & phenotype data management toolkit for Europe

#### Priority in terms of target audience

The learning path was targeted to researchers (biologists) and to data managers specialising in plant data. This targeted audience corresponds to the most urgent needs of the community in terms of capacity building.



**Figure 1** : Learning path for plant phenotyping and plant genomic data management targeting researchers and data managers specialised in plant research

(<https://tess.elixir-europe.org/workflows/plant-phenotyping-and-plant-genomic-data-management> )

#### Development of the Plant Science Learning path

Based on this work a learning path has been set for plant phenotyping and plant genomic data management (Figure 1) with 4 domains: FAIR essentials, plant biological material, plant phenotyping experiment and omics experiments. Within these 4 domains, different thematics are addressed, the “FAIR essentials” domain is an introduction to all following courses. It presents how research data can and should be FAIR and where to publish these data. The “plant biological material” domain is focused

on the biological material metadata that are specific to the plant research domain, essential for interoperability of the data sets and for their re-use. The “plant phenotyping experiment” domain is centred on the optimal way to design and describe phenotyping experiments: it is also very specific to the plant research domain. And finally the “omics experiments” domain is focused on metadata and standards used for each category of omics experiments: the corresponding metadata and guidelines for data management are mostly generic (except the description of the samples).

The learning path includes sub-learning paths for each domain, but the entire set of learning outcomes provides a solution for the integration of plant data. At the end of this learning path, researchers or data managers specialised in plant research are able to:

- retrieve metadata for the biological materials of their interest
- retrieve unique identifiers for their samples from BioSamples
- retrieve MIAPPE compliant metadata on phenotyping experiments
- publish MIAPPE compliant metadata on phenotyping experiments in the recommended databases
- create new ontologies/vocabularies on CropOntology for their projects
- retrieve metadata on omics experiments
- publish metadata on omics experiments in the recommended databases
- use appropriate tools to version their metadata curation before publication

### **Mapping existing training resources on the Plant science learning path**

In order to facilitate the mapping of existing training materials on the learning path, a card has been created for each training, an example is presented in Figure 2 for the “Hands on plant phenotyping experiment metadata : MIAPPE standard” course.

P2 : Hands on plant phenotyping experiment metadata : MIAPPE standard	
URL - TeSS	Not in TeSS
URL - original	<a href="https://github.com/MIAPPE/training/tree/master/Paris-Moulon-Sep-2021">https://github.com/MIAPPE/training/tree/master/Paris-Moulon-Sep-2021</a>
Title/description	Detailed description of the process for collecting metadata of a plant phenotyping experiment.
Authors / Contributors	Cyril Pommier, Daniel Faria, Célia Michotey, Elizabeth Arnaud, Anne Françoise Adam Blondon
Date last updates / Status	Sept 23, 2021
Type	Slides (pptx/pdf)
Learning path	
Prerequisites	NA
Learning objectives	<ul style="list-style-type: none"> <li>- Get acquainted with plant phenotyping data management constraints and particularities</li> <li>- Be able to select and use relevant data standards</li> </ul>

Expected learning objectives	Retrieve a MIAPPE compliant set of metadata of their own use case of plant phenotyping experiment
FAIR-related info	
Availability of materials	In GitHub
Associated code and data	Yes
FAIR evaluation	8/17 <ul style="list-style-type: none"> <li>- Title → Yes</li> <li>- Contact details → No</li> <li>- Description → Yes</li> <li>- Keyword → No</li> <li>- Target audience → No</li> <li>- Learning outcomes → Yes</li> <li>- Prerequisites → No</li> <li>- Persistent identifier → Yes</li> <li>- Availability in a repository → Yes (GitHub)</li> <li>- Availability in TeSS → No</li> <li>- Licensing and (re)use details → No</li> <li>- Preferred citation → No</li> <li>- Required resources → No</li> <li>- Structure and duration → Yes</li> <li>- Additional information → No</li> <li>- Links and references → Yes</li> <li>- Data of last revision → Yes</li> </ul>

**Figure 2.** Card summarising the review of the existing training materials on the topic “Hands on plant phenotyping experiment metadata : MIAPPE standard”

The mapping of existing training materials against the “Plant Science” learning path is summarised in Table 3 together with their range in terms of FAIR score.

**Table 3.** Summary table of the training materials identified for the learning path. The topics in bold are the areas where no training material could be found, see appendix for references on the material found. The FAIR score is given as a range from the lowest to the highest score on the 17 criteria mentioned above.

Training	No. of training found	FAIR score
Basics of the FAIR principles	4	[12-14]/17
Focus on identifiers	3	[8-11]/17
Focus on ontologies and metadata	2	13/17
FAIRsharing, bio.tools, <b>re3data</b>	2 (FAIRsharing and Bio.tools)	[4-13]/17
How to publish your data	1	12/17
<b>e!DAL, recherche.data.gouv, BioData, Zenodo, dataverse</b>	2 (e!DAL and all the tools)	[7.5-15]/17
Introduction to MCPD and MIAPPE	3	[5-12]/17



Training	No. of training found	FAIR score
Hands on collection of plant material and sample metadata	0	
BioSamples	1	6.5/17
Introduction to MIAPPE, ISA, CropOntology	1 (CropOntology)	5.5/17
Hands on plant phenotyping experiment metadata : MIAPPE standard	1	8/17
Plant phenotyping related ontologies and vocabularies : CropOntology, Plant Ontology, ...	1	7/17
PHIS, COPO, FAIRDOME-SEEK, MIAPPE@dataverse (recherche.data.gouv, BioData)	1 (FAIRDOME-SEEK)	6.5/17
Introduction to existing omics metadata standards	0	
Hands on omics metadata and standards : sequencing, transcriptomics,...	0	
ENA, EVA, PRIDE, MetaboLight, Bioproject, data brokering tools	6	[4.5-8]/17

### Identification of gaps

The gaps have two origins:

- Some topics are not covered as well as others in terms of training material.
- The material of some courses is not publicly available and therefore not reusable.

As illustrated in the table above, training material has not been identified for many tools many of them are self-explanatory when it comes to use them (bio.tools, re3data, recherche.data.gouv, Zenodo; dataverse) but others might require trainings (e.g. PHIS; COPO). Larger topics, e.g. how to publish your data; hands on collection of plant material and sample metadata, introduction to existing omics metadata standards, will also have to be addressed.

In terms of FAIR scores, ratings fluctuate across topics, with training materials related to the FAIR essential area rated significantly better than others. For the majority of events that resulted in the creation of training materials, the training materials are found and reusable. However, the repository might not be durable in the long term (e.g. google drive) and when the event is not registered in TESS, there is no unique identifier associated with the deposition, which is important for citation. On the FAIR criteria used for the FAIR score, some criteria are always mentioned such as the title or the links and references. Others are almost never mentioned, such as the resources required to complete the training or the structure and duration of the training.

Complete references for training materials that were used for creating the learning paths are listed in Table 4.



**Table 4.** References of training materials for each learning path topic.

Training	Reference
Basics of the FAIR principles	<a href="https://zenodo.org/record/4629719">https://zenodo.org/record/4629719</a> <a href="https://zenodo.org/record/5078286">https://zenodo.org/record/5078286</a> <a href="https://zenodo.org/record/6561244">https://zenodo.org/record/6561244</a> <a href="https://docs.google.com/document/d/1KXMawj2VsNMD4vqKFEnWD32igpSGkSNuIB_OtLHtHNpM/edit#heading=h.m1kv3qf69cf7">https://docs.google.com/document/d/1KXMawj2VsNMD4vqKFEnWD32igpSGkSNuIB_OtLHtHNpM/edit#heading=h.m1kv3qf69cf7</a>
Focus on identifiers	<a href="https://github.com/OHSUBD2K/BDK11-Identifiers/blob/master/BDK11_01.pptx">https://github.com/OHSUBD2K/BDK11-Identifiers/blob/master/BDK11_01.pptx</a>
Focus on ontologies and metadata	<a href="https://zenodo.org/record/6561286">https://zenodo.org/record/6561286</a> <a href="https://docs.google.com/document/d/1KXMawj2VsNMD4vqKFEnWD32igpSGkSNuIB_OtLHtHNpM/edit#heading=h.ur1feafhs1xx">https://docs.google.com/document/d/1KXMawj2VsNMD4vqKFEnWD32igpSGkSNuIB_OtLHtHNpM/edit#heading=h.ur1feafhs1xx</a> and <a href="https://docs.google.com/document/d/1KXMawj2VsNMD4vqKFEnWD32igpSGkSNuIB_OtLHtHNpM/edit#heading=h.knhkpebtm7uu">https://docs.google.com/document/d/1KXMawj2VsNMD4vqKFEnWD32igpSGkSNuIB_OtLHtHNpM/edit#heading=h.knhkpebtm7uu</a>
FAIRsharing	<a href="https://ifb-elixirfr.github.io/IFB-FAIR-data-training/modules/module3.html">https://ifb-elixirfr.github.io/IFB-FAIR-data-training/modules/module3.html</a>
Bio.tools	<a href="https://zenodo.org/record/7024050">https://zenodo.org/record/7024050</a>
How to publish your data	<a href="https://zenodo.org/record/5524794">https://zenodo.org/record/5524794</a>
e!DAL	<a href="https://doi.ipk-gatersleben.de/DOI/ce4cbdd3-ec68-4fc2-8d6d-da0c87be172b/81ee3de1-4995-480c-9f1b-99cce634c36f/2">https://doi.ipk-gatersleben.de/DOI/ce4cbdd3-ec68-4fc2-8d6d-da0c87be172b/81ee3de1-4995-480c-9f1b-99cce634c36f/2</a>
e!DAL, recherche.data.gouv, BioData, Zenodo, dataverse	<a href="https://zenodo.org/record/6561256">https://zenodo.org/record/6561256</a>
Introduction to MCPD and MIAPPE	<a href="https://github.com/MIAPPE/training/tree/master/Paris-Feb-2020">https://github.com/MIAPPE/training/tree/master/Paris-Feb-2020</a> <a href="https://www.youtube.com/watch?v=4FOQPAWI6_M">https://www.youtube.com/watch?v=4FOQPAWI6_M</a> <a href="https://github.com/PBR/elixir-fondue-datathon/tree/master/presentations">https://github.com/PBR/elixir-fondue-datathon/tree/master/presentations</a>
BioSamples	<a href="https://docs.google.com/presentation/d/1I-Kq1Gr4-989MIUt9xtkyIG2_in1gmmd0QBQRD2QhIs/edit">https://docs.google.com/presentation/d/1I-Kq1Gr4-989MIUt9xtkyIG2_in1gmmd0QBQRD2QhIs/edit</a>
CropOntology	<a href="https://bigdata.cgiar.org/blog-post/webinar-crop-ontology-semantic-interoperability-of-breeding-data/">https://bigdata.cgiar.org/blog-post/webinar-crop-ontology-semantic-interoperability-of-breeding-data/</a>
Hands on plant phenotyping experiment metadata : MIAPPE standard	<a href="https://github.com/MIAPPE/training/tree/master/Paris-Moulon-Sep-2021">https://github.com/MIAPPE/training/tree/master/Paris-Moulon-Sep-2021</a>
Plant phenotyping related ontologies and vocabularies : CropOntology, Plant Ontology,...	<a href="https://github.com/MIAPPE/training/blob/master/Paris-Moulon-Sep-2021/4-Crop-ontology.pptx">https://github.com/MIAPPE/training/blob/master/Paris-Moulon-Sep-2021/4-Crop-ontology.pptx</a>
ENA	<a href="https://raw.githubusercontent.com/jtomasmatus/methada2020/master/other_presentations/metadata2.pdf">https://raw.githubusercontent.com/jtomasmatus/methada2020/master/other_presentations/metadata2.pdf</a> <a href="https://drive.google.com/file/d/12kONh3kVeRshuRBA7VSE6Qo5yWGY174C/view">https://drive.google.com/file/d/12kONh3kVeRshuRBA7VSE6Qo5yWGY174C/view</a>
EVA	<a href="https://docs.google.com/presentation/d/1W8va87XGc2mKwWCOMCIFPQI0JI05BUx_HyuWh00f0Sk/edit">https://docs.google.com/presentation/d/1W8va87XGc2mKwWCOMCIFPQI0JI05BUx_HyuWh00f0Sk/edit</a>

Training	Reference
MetaboLight	<a href="https://integrape.eu/wp-content/uploads/2021/10/Payne_MetaboOPEN_TrainingSchool.pdf">https://integrape.eu/wp-content/uploads/2021/10/Payne_MetaboOPEN_TrainingSchool.pdf</a> <a href="https://drive.google.com/file/d/1uKCGXXe6bKefo2QbbXLI4mkSoMwGFvdM/view">https://drive.google.com/file/d/1uKCGXXe6bKefo2QbbXLI4mkSoMwGFvdM/view</a>

## 6. Conclusions

The method set up by the ELIXIR Training Platform to better structure the development of training resources was proved very helpful to specify the needs in terms of training and capacity building in the context of the use cases. Spelling out the priority in terms of training target and the learning path to follow to reach this target is a way to detail the different modules that are necessary. It is then possible to map on these modules the training resources that are already available and to derive a list of gaps. This method was applied for three demo cases successfully in the Plant, Marine metagenomics and Human Genome data contexts.

In terms of general organisation of community resources to facilitate this type of approach, the most important actions are to make the materials accessible for reuse and easy to find. Accessibility of training materials collected in the project was often fulfilled through non persistent/not versioning file systems and without the association of a persistent identifier. Versioning is highly desirable, to facilitate collaborative development and maintenance of training material of the time and a DOI attribution will facilitate citation and long term findability. Several solutions can be explored including at the national level. At the transnational level, [Zenodo](https://zenodo.org/)<sup>10</sup> is an agnostic repository that delivers DOI and allows versioning. Findability requires additional work to organise the training events and training resources in categories and/or facilitate filtering in the context of catalogues of training resources such as [TeSS](https://teess.elixir-europe.org/)<sup>11</sup> or [Glittr](http://glittr.org)<sup>12</sup> or of knowledge hubs such as the [RDMkit](https://rdmkit.elixir-europe.org/)<sup>13</sup>. This is achieved in TeSS, through the annotation of the training material and linked training events with ontology terms from EDAM, as well as with custom keywords. These annotations also facilitate automatic queries of TeSS by the RDMkit to feed “training” window of its pages and gives more visibility to events and materials in the two portals. “Your domain” RDMkit pages are a way to categorise domain-specific materials produced by ELIXIR members or other providers.

## 7. Impact

The activities carried out in order to identify the needs in terms of training and capacity building for FAIR data management in the context of several use cases have allowed to test a method aiming at structuring the description of these needs that can be now be re-used on other domains of applications or use cases either in transnational projects (e.g. it has been implemented by the [ELIXIR System Biology community](https://systembiologycommunity.org/)<sup>14</sup> and is planned to be implemented in the newly formed ELIXIR Biodiversity community) or within nodes (e.g. the French INRAE data steward community has started

<sup>10</sup><https://zenodo.org/>

<sup>11</sup><https://teess.elixir-europe.org/>

<sup>12</sup><http://glittr.org>

<sup>13</sup><https://rdmkit.elixir-europe.org/>

<sup>14</sup><https://teess.elixir-europe.org/workflows/building-interoperable-and-reusable-systems-biology-models>

to develop the training material necessary to implement the “Plant phenotyping experiment” domain of the Plant use case learning path).

The work done on the FAIRness of the training materials has stressed two very important features for collaborative work and therefore for capacity building: access and re-use of training material, which necessitate sharing training material on long term sustained open platforms that allow versioning and attach a DOI. Another important feature is the findability of the material and the methods and resources necessary for increasing findability of training materials in TeSS or in the RDMkit have been set up, including a collaboration with the EDAM ontology.

## 8. Next Steps

The perspectives are now to disseminate the work done in ELIXIR’s communities, including the emerging “Data management community” to get feedback and to engage new people in further developing resources addressing the gaps identified.

## 9. Deviation from Description of Action

None

## Annexes

### Annex 1 - Training and capacity building needs survey

#### Section 1: Contact details

Firstname

Lastname

Mobile

Email

---

Q1: What is your job title?

---

Q2: Which is your country of employment?

---

Q3: How would you describe the organization in which you currently work?

- Academia / Research Institution
- Industry / Small and Medium Enterprise (SME)



- Non-Profit Organisation
- Healthcare
- Other

---

Q4: Please select your ELIXIR Node.

---

## Section 2: Training courses and materials

---

Q5: Which is your field of expertise in data management?

---

Q6: Which is (in your opinion) the most important stage of the data management (DM) life cycle?

- Planning research
  - Collecting and creating data
  - Processing and analyzing data
  - Publishing and sharing data
  - Preserving data
  - Re-using data
  - Compliance with regulations (e.g. GDPR)
- 

*Further questions related to the DM cycle will be asked during 1-2-1 talks*

Q7: Which CONVERGE WP5 Demonstrator Project are you involved in:

- Harmonised FAIR plant genotype & phenotype data management toolkit for Europe.
  - Reproducible, comparable and FAIR epitranscriptomics.
  - Common data management plans for the Marine Metagenomics Community.
  - FAIR encoding and access to toxicology data.
  - FAIR organisation of biomolecular simulation information.
  - Federated access to human genomics data: GDPR
  - Other related CONVERGE WPs: CONVERGE WP7 - Federated European Genome-phenome Archives for transnational access of COVID-19 host data
  - Other related CONVERGE WPs: CONVERGE WP9 - Mobilisation of SARS-CoV-2 variant surveillance data tracking services and tools
  - Other CONVERGE WPs: below please write your CONVERGE WP.
  - I am not involved in any CONVERGE WPs. Below please describe your field of expertise.
- 

Please write your CONVERGE WP if you answered "Other CONVERGE WPs" in previous question

---

Please describe your field of expertise if you answered "I am not involved in any CONVERGE WPs" in previous question.

---

Q8: Please indicate the availability of training courses related to your Demonstrator Project / WP / field of expertise.

---



- There are sufficient courses about this topic.
- There are limited courses about this topic.
- To my knowledge, there are no training courses on this topic.

---

Q9: Please provide information about the training courses related to your Demonstrator Project / WP / field of expertise that you are aware of (course title, link to the course, short description, organizer etc.).

---

Q10: Please indicate the availability of training materials related to your Demonstrator Project / WP / field of expertise (materials that are available without attending a course).

- There are sufficient materials about this topic.
- There are limited materials about this topic.
- To my knowledge, there are no training materials on this topic.

---

Q11: Please provide information about training materials related to your Demonstrator Project / WP / field of expertise that you are aware of (title and format of the material, link to the material, short description etc.).

---

Q12: Are you willing to participate as a trainer at a future ELIXIR course covering your Demonstrator Project / WP / field of expertise?

- Yes
- Maybe.
- Not personally but I can recommend someone. Below please name the person(s) that you recommend and add their affiliations.
- No.

---

Q12: Please name the person(s) that you recommend and add their affiliations if you chose "Not personally but I can recommend someone" in previous question.

---

Q13: Are you willing to author training materials for a future ELIXIR course covering your Demonstrator Project / WP / field of expertise?

- Yes.
- Maybe.
- Not personally but I can recommend someone (please name her / him). Below please name the person(s) that you recommend and add their affiliations.
- No.

---

Q13: Please name the person(s) that you recommend and add their affiliations if you chose "Not personally but I can recommend someone" in the previous question.

---

Q14: Please name one or more topics related to your Demonstrator Project / WP / field of expertise where (in your opinion) there is a lack of topic-specific data management training. If you name more than one topic, please list the topics in the descending order of importance.

---

Q15: Please name one or more topics outside your Demonstrator Project / WP / field of expertise where (in your opinion) there is a lack of topic-specific data management training. If you name more than one topic, please list the topics in the descending order of importance.

---

### **Section 3: Follow-up and final comments**

---

Q16: Do you have any other comments?



## Annex 2 - Inventory of training materials in the context of the use cases

### Summary of courses listed in WP5 training inventory

Course title	Contact person	ELIXIR Node(s)	Date (DD/MM/YYYY format)
Plant Phenotyping Data management (MIAPPE)	Cyril Pommier	FR, NL, PT	17/02/2020 - 18/02/2020
Plant Phenotyping Data management Webinar (MIAPPE)	Cyril Pommier	FR	17/12/2020
Data protection in research (practical with DAISY)	Marina Popleteeva	LU	20/05/2020
Methada 2020	Tomas Matus	ES	05/02/2020 - 07/02/2020
Methada 2020 eHands-on	Jerome Grimplet	ES	30/11/2020 - 03/12/2020
METABO-OPEN 2021	Fulvio Mattivi	IT	04/10/2021 - 06/10/2021
METABO-OPEN 2022	Fulvio Mattivi	IT	15/06/2022 - 17/06/2022
Annot'Training 2022	Camille Rustenholz	FR	28/06/2022 - 29/06/2022
ELIXIR FONDUE Datathon	Sebastian Beier	DE,NL,FR,EMBL-EBI	15/06/2021 - 17/06/2021
Functional gene ontologies for plants and their use in data integration	Kristina Gruden	SI	21/04/2020



Project data management using pISA-tree	Maja Zagorščak	SI	19/03/2021
GCBN de.NBI user training - FAIR data management for plant genomics and phenomics	Daniel Arend	DE	13/06/2018
The de.NBI/ELIXIR service "PGP" for phenomics data publication	Daniel Arend	DE	asynchronous
Best practice recommendations through exemplar Data Management Plans in the Data Stewardship Wizard	Korbinian Bösl, Espen Åberg, Federico Bianchini	NO	06/04/2022
Using the Norwegian e-infrastructure for Life Science and usegalaxy.no	Erik Hjerde, Espen Åberg	NO	23/03/2022, 07/04/2022, 13/04/2022
Data Management Planning workshop for new Life Science Projects	Korbinian Bösl, Espen Åberg, Federico Bianchini	NO	04/05/2020 - 06/05/2020
Hands-on workshop in Marine Metagenomics	Erik Hjerde, Espen Åberg	NO	26/11/2018 - 30/11/2018

