Project acronym: EOSC4CANCER
Grant Agreement Number: 101058427
Project full title: A European-wide foundation to accelerate Data-driven Cancer Research
Call identifier: HORIZON-INFRA-2021-EOSC-01

# D5.1 A guidance document for the EOSC4Cancer learning pathway

| | |
|---|---|
| Version: | 1.0 |
| Status: | Final |
| Dissemination Level: | Public |
| Deliverable Type: | *Report* |
| Due date of deliverable: | 30.11.2023 |
| Actual submission date: | 23.11.2023 |
| Work Package: | WP 5 Training and capacity building for key national and international stakeholders |
| Lead partner for this deliverable: | CERTH |
| Partner(s) contributing: | HEALTH RI, CNIO, EATRIS, UNIMAN, BBMRI |

Main author(s):

| | |
|---|---|
| Fotis Psomopoulos | INAB\|CERTH |
| Fatima Al-Shahrour | CNIO |

Other author(s):

| | | | |
|---|---|---|---|
| Sarah Morgan | EATRIS | Celia van Gelder | Health-RI/DTL |
| Munazah Andrabi | UNIMAN | Kurt Majcen | BBMRI |
| Fieke Schoots | Health-RI/DTL | | |

**Abstract**

The EOSC4Cancer initiative addresses the pressing need for advanced research and infrastructure in tackling cancer, particularly in Europe. This document focuses on the WP5 initiative within EOSC4Cancer, which aims to empower clinicians and cancer researchers through capacity building. Emphasizing the importance of tailored training, the document outlines the design process for learning paths, including short-term courses and traditional postgraduate options. Key concepts such as learning outcomes and FAIR training are introduced, and the systematic mapping of skills and roles within the target audience is detailed. Illustrative examples from both within and beyond the project demonstrate the application of these principles. The document concludes with a list of practical tools and services for constructing effective learning paths. Importantly, this document is a living resource, reflecting ongoing efforts and intended to contribute to a dynamic learning environment beyond the EOSC4Cancer project, benefiting the broader scientific community.

Funded by the European Union

## Revision History

| Version | Date | Changes made | Author(s) |
|---|---|---|---|
| 0.1 | 10.30.2023 | Initial draft | Fotis Psomopoulos (INAB\|CERTH), Fátima Al-Shahrour (CNIO) |
| 0.4 | 11.03.2023 | Review of the document, addition to the collections | Sarah Morgan (EATRIS), Kurt Majcen (BBMRI) |
| 0.5 | 11.08.2023 | Updates on the "Designing Learning Paths" section | Jessica Lindvall (NBIS), Loredana Le Pera (CNR), Alexia Cardona (Un. Cambridge) |
| 0.7 | 11.09.2023 | Updates across introduction and conclusions, revisions in the main text | Fotis Psomopoulos (INAB\|CERTH), Fieke Schoots (HealthRI-DTL) |
| 0.9 | 11.10.2023 | Added the Collation of training resources section, added the TeSS section. | Sarah Morgan (EATRIS), Munazah Andrabi (UniMan) |
| 1.0 | 11.17.2023 | Internal review | Eva Alloza (BSC) |
| | | | |

**Statement of originality**

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

# TABLE OF CONTENTS



# LIST OF FIGURES

## LIST OF TABLES

# EXECUTIVE SUMMARY


## Introduction

Cancer, a complex and pervasive disease affecting individuals worldwide, demands a concerted effort for prevention, diagnosis, and treatment. With Europe shouldering a disproportionate cancer burden, the need for advanced research and infrastructure is paramount. The EOSC4Cancer initiative seeks to leverage existing resources and community solutions, establishing a federated FAIR data, analysis, and services infrastructure dedicated to European Cancer research programs. In the face of escalating data complexity and technological advancements, effective analysis and interpretation of cancer research data have emerged as critical components for understanding the disease and enhancing treatment outcomes. The WP5 initiative within EOSC4Cancer is committed to empowering clinicians and cancer researchers with the requisite skills and knowledge to leverage the project's tools and infrastructure effectively, both during and beyond the project lifecycle.

## Designing Learning Paths

Recognizing the significance of capacity building, this document underscores the importance of designing tailored training solutions in the form of structured learning paths. Emphasising the efficacy of short-term courses, typically spanning a week, in providing comprehensive insights and hands-on experiences, it also acknowledges the value of traditional postgraduate and specialisation courses in this domain. The document outlines the learning path design process, after providing sufficient context in key underlying concepts such as learning outcomes and FAIR training. The process itself comprises a systematic mapping of the requisite skills and pertinent roles within the target audience (the EOSC4Cancer stakeholders in this instance), using these as a driver towards identifying the necessary topics. The application of these guiding principles for constructing learning paths are demonstrated through illustrative examples derived both from within the project and beyond. Finally, the document includes a short list of practical tools and services conducive to the seamless construction of effective learning paths.

## A Living Document

This document represents a comprehensive consolidation of efforts in shaping learning paths within the EOSC4Cancer framework. It is pertinent to note that the document's content reflects ongoing initiatives, anticipated to extend beyond the EOSC4Cancer project, fostering a dynamic learning environment for the benefit of the broader scientific community.

# 1 Introduction

Cancer is a heterogeneous disease that affects almost everyone as patient, survivor, relative or friend. Cancer is also a major societal issue with Europe having more than 25% of the cancer cases for only 10% of the world population, a situation that will deteriorate further in our ageing societies where 13 million patients will be considered long-term cancer survivors in 2035. Developing cancer prevention, diagnosis and treatment is complex: the individuals' genetic composition, behaviour and environment all interact to determine risk and treatment outcomes. EOSC4Cancer builds on existing projects, research outcomes and established community solutions to create the federated FAIR data, analysis and services infrastructure needed for European Cancer research programmes.

Given the exponential growth in data complexity and the rapid advancements in technology, the effective analysis and interpretation of cancer research data have become critical for advancing our understanding of the disease and improving treatment outcomes. However, this pursuit is accompanied by various challenges, including the need to equip researchers with the necessary skills, facilitate seamless knowledge transfer, and foster a collaborative environment that promotes innovation and excellence. As such, WP5 in EOSC4Cancer aims to ensure that the clinicians and cancer researchers will have the knowledge and skills on how to best address their scientific questions by using the tools and infrastructure developed in the lifecycle of the project and beyond.

In the vast majority of capacity building activities, the optimal format to design and deliver training are short-term courses. These activities, usually lasting up to a week, can provide enough content to introduce participants to an overview of the necessary skills and hands-on training experience that will allow them to take full advantage of the relevant tools and services. Of course, these short-term activities are complemented by the traditional training courses that can be found in the context of postgraduate training programmes and specialization courses.

However, a common theme in both cases is the requirement to formally structure, design and ultimately deliver the training, in a way that (a) fully matches the expected skills that need to be acquired, and (b) allow for effective reuse of the content. This document aims to provide guidance and practical examples on how to achieve this - although defined in the context of the EOSC4Cancer project, it’s expected to be transferable in other domains as well.

The document is structured as follows; Section 2 lists the key definitions that will be used throughout the document, as well as the foundational principles in designing training resources. Section 3 focuses on the mapping of skills and relevant roles in the EOSC4Cancer project, which will act as the basis for identifying the necessary courses. Section 4 will provide the guiding principles in designing learning paths, including some key examples from both within the project and beyond. Finally, Section 5 provides some practical services that can be effectively used to design learning paths.

It should be noted that this document is a collection of all the information and efforts around the design of Learning Paths. A lot of these activities are still ongoing and will be used as part of the EOSC4Cancer design process as well beyond the date of this document.

# 2 Key Concepts

## 2.1 Key definitions

The following definitions have been adapted from relevant works[1]:

| Type | Categories |
|---|---|
| Learning area | A learning area is defined as the "*Grouping of traditionally discrete but related subjects with the explicit aim of integrating students' learning*"[2]. |
| Learning topic | In this document, we define learning topics as teaching units that include a presentation of the theory (corresponding to a learning outcome) and one or several learning activities (e.g., an exercise corresponding to a learning outcome). A topic can usually cover two or fewer hours of teaching. |
| Learning outcome | "*Totality of information, knowledge, understanding, attitudes, values, skills, competencies or behaviours an individual is expected to master upon successful completion of an educational programme*"[3]. |
| Learning activity | The specific activity that will support learners to achieve the learning outcomes. Learning activities are coupled with specific learning outcomes at each level of the Bloom's taxonomy[4]. They can refer both to theoretical lectures and to practical exercises. |

**Table 1**: Definitions of Learning Concepts

## 2.2 Learning Paths

> **Learning path**. Selection and interconnection of learning topics tied together for learners to progress through, while mastering a particular subject or program[5].

Learning Paths are an important support for any learner in their professional development trajectory. Today, many disciplines, domains and competencies have designed dedicated training collections and linked learning paths to them as to assist individuals in what

[1] Anteghini, M., Elez, K., Bandstra, S., Lamba, R., & Paladin, L. (2023, June 15). Bioinforming. https://doi.org/10.37044/osf.io/p8n2t

[2] UNESCO International Bureau of Education. (2013). Glossary of curriculum ter-minology. https://unesdoc.unesco.org/ark:/48223/pf0000223059.locale=en

[3] UNESCO Institute for Statistics. (2012). International standard classification of education: ISCED 2011. https://uis.unesco.org/sites/default/files/documents/international-standard-classification-of-education-isced-2011-en.pdf

[4] Bloom, B. S. (1956). Taxonomy of educational objectives: The classification of educational goals. Handbook 1, cognitive domain. David McKay

[5] Nabizadeh, A. H., Leal, J. P., Rafsanjani, H. N., & Shah, R. R. (2020). Learning path personalization and recommendation methods: A survey of the state-of-the-art. ExpertSystems with Applications,159, 113596. https://doi.org/10.1016/j.eswa.2020.113596

"Knowledge, Skills and Behaviours (or attitudes)" areas they need to train themselves in, in order to become proficient or expert in a certain domain. Challenges however do persist in designing such learning paths, both for the use from the Trainer/Instructor perspective as well as from the learner perspective.

A learning path is a pathway that guides the learner through a set of learning modules (courses/materials) to be undertaken progressively, from lower- to higher-order thinking skills, to acquire the desired knowledge and skills on a subject by the end of the pathway. In practice, learning paths help educators (a) design effective educational programmes, and (b) reduce the gaps in their curriculum.

A learning path needs to have three key properties; be progressive, have a direction, and have a defined start and end point.

## 2.3 FAIR Training

A major task in developing new training sessions is preparing training materials, which can be time consuming and challenging for both new and experienced trainers. One solution is to find and (re)use existing materials. This requires that they have been shared, properly described, and made available for (re)use by their authors; but finding suitable online materials that aren't subject to licensing and/or copyright restrictions can be hard. They are also often scattered across different repositories, are siloed in their home institutions, or lack the metadata required to enable their (re)use. If we are to meet the demand for bioinformatics and computational biology training, we need to share and deliver training materials consistently, following best practices that enable their (re)use and adaptation. The scientific community has attempted to address this issue by developing a set of rules to help make training materials easier to find, (re)use, and adapt, for the benefit of all.[6]

In summary, the rules are:

- Rule 1: Plan to share your training materials online
- Rule 2: Improve findability of your training materials by properly describing them
- Rule 3: Give your training materials a unique identity
- Rule 4: Register your training materials online
- Rule 5: Define access rules for your training materials
- Rule 6: Use an interoperable format for your training materials
- Rule 7: Make your training materials (re)usable for trainers
- Rule 8: Make your training materials usable for trainees
- Rule 9: Make your training materials contribution friendly
- Rule 10: Keep your training materials up-to-date

Making training material FAIR by design is key in ensuring both high impact and effectiveness in training, and has been recognized across a number of activities and projects. It is worth highlighting the Skills4EOSC project, which has recently produced the FAIRbyDesign Methodology for learning material[7] (**Figure 1**).

[6] Garcia L, Batut B, Burke ML, Kuzak M, Psomopoulos F, et al. (2020) Ten simple rules for making training materials FAIR. PLOS Computational Biology 16(5): e1007854. https://doi.org/10.1371/journal.pcbi.1007854

[7] Filiposka, S. (2023). D2.2 Methodology for FAIR-by-Design Training Materials (1.4). Zenodo. https://doi.org/10.5281/zenodo.8305540


**Figure 1**: The FAIR-by-Design methodology of the Skills4EOSC project

# 3 EOSC4Cancer skills and topics

In order to effectively create Learning Paths, it is important to have a clear mapping of the involved skills and the respective topics that will be tackled.

## 3.1 Training actors and stakeholders

In their 2021 report on "Digital skills for FAIR and open science"[8], the EOSC executive board skills and training working group set out a framework of EOSC actors - defining roles that would need to interact with EOSC, and the form those interactions might take. Under deliverable 6.1 for EOSC4Cancer, these roles were expanded under a cancer lens, thereby providing a more focused view of their skills needs within the cancer landscape.

| EOSC Ecosystem Role | Description | Examples in EOSC4Cancer |
|---|---|---|
| Researcher | Main focus of the EOSC ecosystem, accessing it to obtain, process, create, store and share research data. | Cancer researchers: would browse EOSC4Cancer data sets to carry out a study |
| EOSC Enabler | Researcher with both profound domain knowledge and open science and FAIR skills, able to design discipline-specific applications with EOSC services and data. Acting as a bridge between scientific community and technical services. | Cancer researchers with technical skills who integrate genomic data processing suite into EOSC4Cancer infrastructure (e.g., user in open science initiatives like Galaxy, cBioPortal etc.) |
| Data Scientist/ Data Analyst | Data processing expert who can evaluate data quality and extract and represent relevant knowledge | Data scientist developing a machine learning algorithm for the EOSC4Cancer federated learning infrastructure |
| Research Software Engineer | IT expert designing, implementing, and maintaining software in the EOSC ecosystem | Software engineer designing software for EOSC4Cancer |
| Data Research Infrastructure Support Professional | ICT expert managing and operating research infrastructures and services to store, preserve and process research data | Professional managing and operating research infrastructures in EOSC4Cancer |
| EOSC Educator | Expert organising content and managing training activities in EOSC. Thorough knowledge of the EOSC | EOSC4Cancer training organisers on institutional or national level |

[8] European Commission, Directorate-General for Research and Innovation, Manola, N., Lazzeri, E., Barker, M. et al., Digital skills for FAIR and Open Science – Report from the EOSC Executive Board Skills and Training Working Group, Manola, N.(editor), Lazzeri, E.(editor), Barker, M.(editor), Kuchma, I.(editor), Gaillard, V.(editor), Stoy, L.(editor), Publications Office, 2021, https://data.europa.eu/doi/10.2777/59065

| | | |
|---|---|---|
| | ecosystem. | |
| Data Curator | Expert managing and overseeing an organisation's entire data to ensure compliance and provide users with high quality, accessible data. | Senior level cancer registry representative – e.g. collecting and publishing genomic data samples related to cancer, with geographic annotation, in domain-specific standard formats |
| Data Steward/ Data Librarian | Expert for preparing and treating data, including selection, storage, preservation, and metadata maintenance. | Expert adding metadata to a genomic dataset related to cancer |
| Citizen | People interested in science, wishing to contribute to citizen science initiatives. | Citizens, patients, and survivors anonymously sharing cancer data to EOSC4Cancer |
| Policy Maker | Policy makers extract information from research to build policies or strategic frameworks | Governing board of a public health directorate using EOSC4Cancer data to help implement national cancer task force. |

**Table 2**: EOSC4Cancer actors, as mapped across the EOSC Actor Ecosystem[9]

Given the remit of EOSC4Cancer, including its aims and key outputs for research, the following roles were identified as those which required more specific focus in terms of skills and knowledge acquisition.

- Pre-clinical cancer researcher
- Clinical cancer researcher
- EOSC enabler
- Data scientist/analyst
- Research software engineer
- Data RI support professional
- Data curator
- Data steward

It must be noted that there is overlap between a number of these roles, which highlights the importance of producing learning paths which are multidisciplinary in nature for this project. The cancer researcher roles have been subdivided in the initial analysis as there are anticipated differences in data acquisition and use between those working in pre-clinical and clinical settings.

[9] https://drive.google.com/drive/folders/1gqI5HOHA7fWn8C0zewkl6Z1qeR61JNOR

## 3.2 Identifying key topics and priorities

To identify the key topics for skills and knowledge acquisition, a number of reference points within the project and the wider EOSC / Open science community were reviewed:

- Overarching project aims and outcomes focused on the data lifecycle and associated project outputs e.g. data platforms, software, workflows
- Specific use cases within the project that focus on different elements of the patient journey
- Awareness and specialist knowledge of EOSC including access and use
- Awareness and knowledge of Open science practices and methods to enable their delivery
- Translation of research into patient benefit

By reviewing these needs, a set of generic skills and knowledge was initially identified under separate sub headings (e.g. "Data management" as shown below), applicable across the range of EOSC actors as identified in section 2. These generic skills were then reviewed from a domain specific viewpoint, ensuring the focus was on cancer data specific needs (e.g. in the case of data management, specific data standards, ontologies and data repositories). The challenge in providing domain / project specific skills was the level of granularity in e.g. naming specific data standards and ontologies. It was decided to apply a pragmatic approach, using terms such as "appropriate standards for cancer data", but noting that when learning paths are produced, more specific resources will be named / applied within the training.

| Overarching topic | Generic skills and knowledge | Domain / project specific skills & knowledge |
|---|---|---|
| **Data management** | Creating a data management plan (DMP) | Awareness of restrictions due to data types (Clinical especially) and appropriate services for DMPs, both for primary and secondary data |
| | Data standards and their use (including metadata) | Applying appropriate data standards / harmonised approaches for data types involved e.g., clinical data standards |
| | Ontologies and their use | Cancer specific ontologies / controlled vocabularies |
| | ELSI for sensitive data use - esp in a cloud context (e.g., GDPR & data security) | National regulations |
| | FAIR — Findable, Accessible, Interoperable, and Reusable — data | Appropriate repositories, tools, standards for cancer data and /or related patient data e.g., EGA |
| | Open science and public sharing of data | Awareness of the data sensitivity, anonymisation and levels of access and confidentiality applicable |
| | Federated data | Specific considerations for cancer data |

**Table 3**: Indicative structure of Topics, across generic and domain-specific skills

## 3.3 Mapping of skills

It is evident that a particular skill might be relevant to multiple roles at different levels of expertise.

In order to facilitate this, we propose the following three levels that will ultimately correspond to the three stages that a Learning Path (LP) might address, covering one or more of them depending on the topic and the constructed path (see section 4.1 below):

- **Level 1: Awareness.** Knowledge of basic principles and surface level understanding of impact of these on their work / work of others.
- **Level 2: Able to use / apply.** More in-depth knowledge of principles and the ability to consistently and appropriately apply the knowledge / skills when required.
- **Level 3: Able to develop / implement**. Need to be fully proficient on this topic, able to support others in applying this knowledge / skill and with the potential to advance the development of the topic. Roles working at this level may be responsible for implementing processes / tools etc so that others can apply them (i.e. supporting those working at Level 2 or below).

Examples of this mapping are listed below:

| Role | Topic | Level of expertise required |
|---|---|---|
| Clinical cancer researcher | Creating a data management plan (DMP), being aware of restrictions due to data types. | Able to use / apply |
| Clinical cancer researcher | Ontologies and their use, Cancer specific ontologies / controlled vocabularies | Awareness |
| Clinical cancer researcher | Data Interpretation to be used in Clinical Decision Support systems | Able to develop / implement |
| Data curator | Creating a data management plan (DMP), being aware of restrictions due to data types. | Able to develop / implement |
| Data curator | Cloud ready workflows, running cancer analysis workflows e.g., Galaxy | N/A |
| Clinical cancer researcher | Cloud ready workflows, running cancer analysis workflows e.g., Galaxy | Able to use / apply |

**Table 4**: Example mapping of skills and roles

It is important to note that having a clear and concise mapping of the skills across the topics should allow for the identification of gaps and areas for improvement in the current skill sets of researchers and stakeholders, and when cross referenced with an available set of training resources, can also provide a view of gaps in training provision.

## 3.4 Collation of training resources

In order to move towards defining learning paths, there is a need to not only define the skills and knowledge that individuals must acquire, but also to signpost training resources that learners are able to access which meet these needs. These resources may be available in a variety of formats e.g. in person or online, synchronous or asynchronous; and could vary in nature from being full short courses to webinars, tutorials, or self-led practicals.

The concept of a curated materials collection, which brings together a themed set of resources which allow for the acquisition of skills / knowledge across a set topic, can be a first step within the learning path development process. These collections of materials can be developed using a range of resources which can provide a start point to a topic, applicable across a range of audiences. These are learning resources in their own right, and have been successfully employed to provide a start point to learning across a range of introductory topics. Though there is logic in their ordering, they are less structured than a learning path (where trainees expect to move from learning point a to b, to c) but are presented in a logical manner whereby topics within a theme link together, but with less building upon knowledge as you move from one resource to another.

A recent example of these curated collections are those developed by the EMBL-EBI training team under the CONVERGE project, focused on Data management and Biocuration. These collections were built from a combination of existing materials (e.g., recorded webinars, online tutorials relating to EMBL-EBI resources, excerpts from short live courses and introductory tutorials) and new materials where required. Each collection provides a set of learning outcomes for the whole collection, along with the intended audience and any required pre-requisite knowledge. A narrative is written to provide links between each learning resource and interactive quizzes and activities to check one's own learning are also interspersed through the collection. Though presented in an ordered manner, each piece of the collection can be accessed individually and the learning undertaken in any order which the learner wishes to, providing a flexible structure.

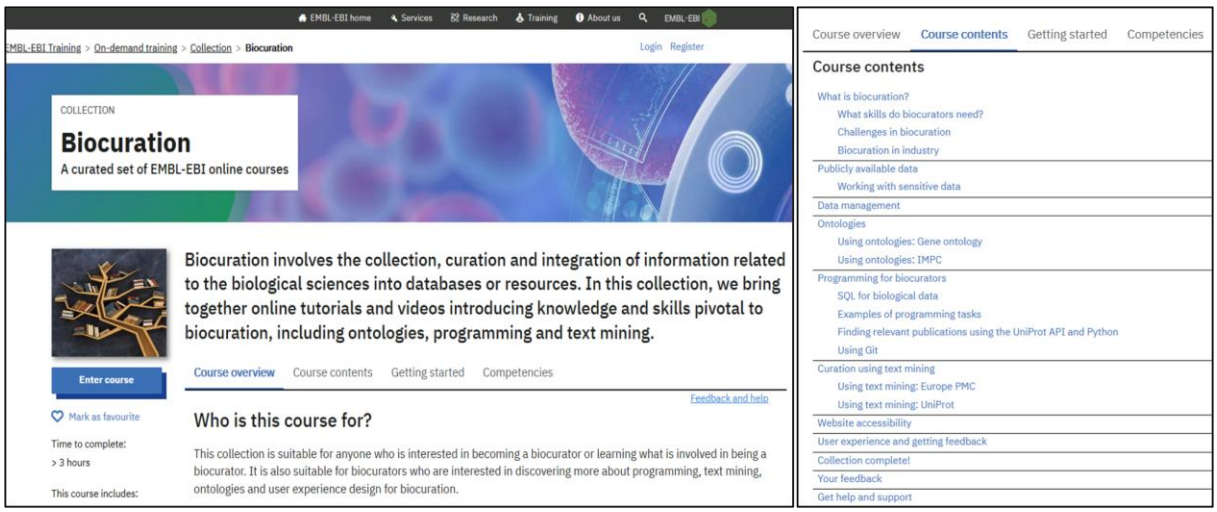


**Figure 2**: Front pages of an example of a curated collection of resources from EMBL-EBI in biocuration, developed under the CONVERGE project.Biocuration | EMBL-EBI Training

Within EOSC4Cancer, we are in the process of collating training resources from across project partners, training and research organisations working in research data and cancer, the EOSC

network and related cancer projects (e.g., CanSERV and EUCAIM). The aim is to collect examples of the range of training resources that are available to upskill researchers, data handlers/analysts and infrastructure/software professionals in the handling of cancer research data and use of EOSC resources. To date this has involved reviewing the types of resources available that are applicable in learning path development, collating information on the organisers and developers of appropriate training resources and collecting specific examples of training that is currently available. There has been a specific focus on gathering training in relation to the specific outputs of EOSC4Cancer, to highlight where gaps exist e.g. in supporting the adoption of EOSC4Cancer resources by the wider community.

This training resource collection continues to be under development, and will provide signposted training resources to enable trainees to work through a learning path and meet their end goals of acquiring the skills and knowledge needed; it will also allow for identification of gaps in training provision which can then be highlighted to expert groups within the wider consortium who can then be supported to develop the required materials. By grouping materials into themed sets, we can also provide some “dip in” resources for individuals who want general upskilling in specific topics, but do not wish to follow a full learning path.

# 4 Designing Learning Paths

## 4.1 Guiding Principles

In order to design a learning path we will be following the guidelines of "The LearningPath step-by-step protocol"[10] (**Figure 3**). The protocol refers to a protocol that was developed across two Biohackathon Europe projects (BH2021, BH2022) led by members of the ELIXIR community, primarily coming from within the Training Platform. The protocol is field-agnostic and provides step-by-step guidelines for curriculum developers and trainers.

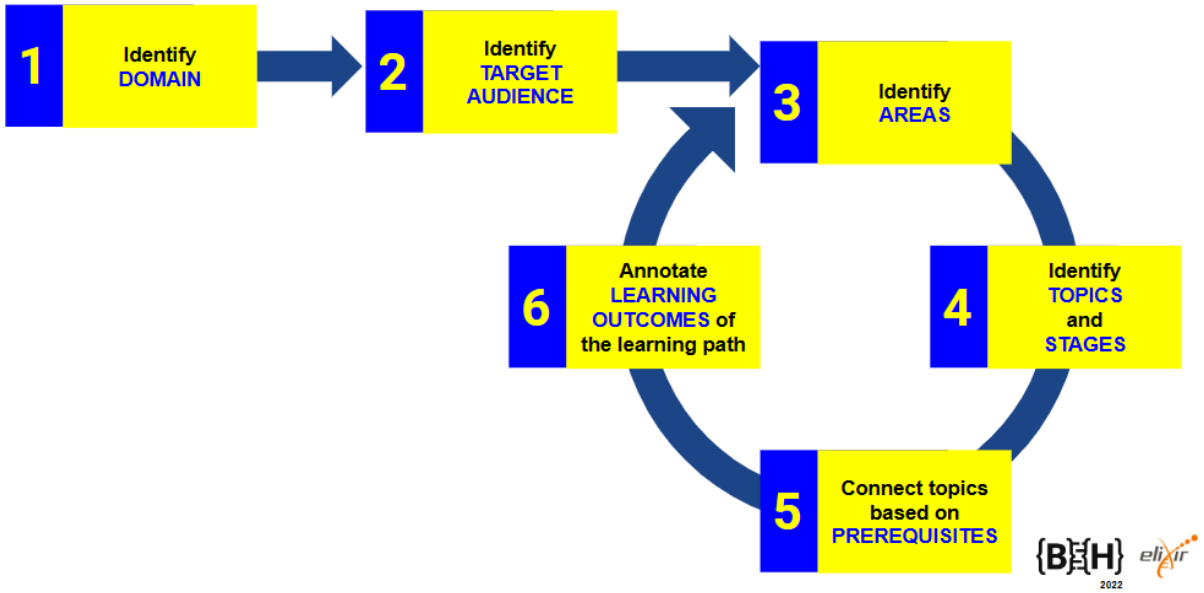


**Figure 3**: The LearningPath step-by-step protocol[10]

In constructing an effective learning path, the initial step involves identifying the domain around which the learning path will revolve. Within the context of cancer research, this would entail focusing on the specific area that necessitates skill development and knowledge enhancement. For instance, the learning path might centre on intricate aspects such as "Statistical Analysis in Cancer Research" or "Bioinformatics in Oncology."

Understanding the target audience is pivotal in tailoring the content and structure of the learning path to meet their specific needs. In the realm of cancer research, the target audience typically comprises a diverse group of professionals, including researchers, clinicians, and data analysts actively engaged in the field of oncology. These individuals usually possess a solid background in oncology research and data analysis, thereby requiring content that aligns with their expertise levels and research objectives.

[10] Cardona, A., Jetten, M., Lindvall, J., & Le Pera, L. (2023, November 7). Developing Learning Paths - The Learning Paths Protocol. Zenodo. https://doi.org/10.5281/zenodo.10082795

Furthermore, the learning path has been meticulously structured into three distinct stages, drawing inspiration from the "*Mastery Rubric for Bioinformatics*" framework[11]. This framework serves as a robust foundation, offering a systematic and evidence-based developmental trajectory, guiding learners toward increased cognitive complexity and proficiency. By amalgamating the Novice and Beginner stages into the Beginner phase, the Intermediate stage encapsulating the Apprentice, and the Advanced stage encompassing the early and late journeyman stages of the MR-Bi framework, we have established a coherent progression pathway for learners within the domain of cancer research data analysis.

Each stage has been thoughtfully crafted, accompanied by comprehensive descriptions that delineate the performance expectations, behaviour patterns, and requisite levels of supervision necessary for learners to effectively navigate and derive maximum benefit from the learning journey. This approach ensures a well-rounded and comprehensive learning experience tailored to the specific needs and expertise levels of the target audience.

| Learning Path Stage | Description |
|---|---|
| Beginner stage | Engages with given problems, with known solutions, learning how to analyse them; doesn't question research results; can use and apply given tools as instructed.<br>(e.g., from early undergraduate to early Master's level) |
| Intermediate stage | Fluent scientist, who can choose and apply methods to given problems, analyse and interpret data, identify basic limitations, and contextualise results; doesn't generate new problems; seeks guidance to improve performance.<br>(e.g., form early PhD to early postdoc level) |
| Advanced stage | From Proficient scientist (synthesises knowledge; beginning to critically evaluate experimental paradigms and their results; accept uncertainty; contributes to problem formulation) to independent scientist (expert in design/critical evaluation of experimental paradigms and their results; expertly integrates bioinformatics into research practice; can apply/develop new methods, formulate problems)<br>(e.g., from Postdoc to Principal Investigator level) |

**Table 5**: The three stages of mastery in the Learning Path design[10]

To help to map the topics in the appropriate part of the framework, we have also added on the left, for each stage, the corresponding cognitive levels, as given in Bloom's taxonomy.

[11] Tractenberg RE, Lindvall JM, Attwood TK, Via A (2019) The Mastery Rubric for Bioinformatics: A tool to support design and evaluation of career-spanning education and training. PLOS ONE 14(11): e0225256. https://doi.org/10.1371/journal.pone.0225256


**Figure 4**: Example template, using areas "Computing", "Statistics/Artificial Intelligence", "Bioinformatics resources", and "Genomics" as the overarching topics[10]

For each area, we need to identify which cognitive level the learner should achieve during the learning path. For example, in the area of "*Computing*", we may map the topic "*R programming*" to the stage "*beginner*" (learners will be able to reach a cognitive level between "remember" and "apply"). The next step would be to map the topics covered during the Learning Pathto the corresponding cognitive level/stage. In order to identify the connections across topics, and for each topic, we should ask: "*Is there any topic in the framework that is a prerequisite for this (it should be taught first)?*". If so, an incoming arrow should be placed connecting the topics together. In the end, we have a picture in which, for each area, we can see the final stage achieved at the end of the Learning Path.

Some notes:

- a "topic" should have at least a couple of hours of content to be taught, otherwise it could be just a concept (too much granularity).
- once a topic is processed, colour the respective box or name in yellow to distinguish it from boxes that have not yet been processed.
- in principle, a "topic" could be a "course". If you have an archive of courses, you could map these courses into the framework.

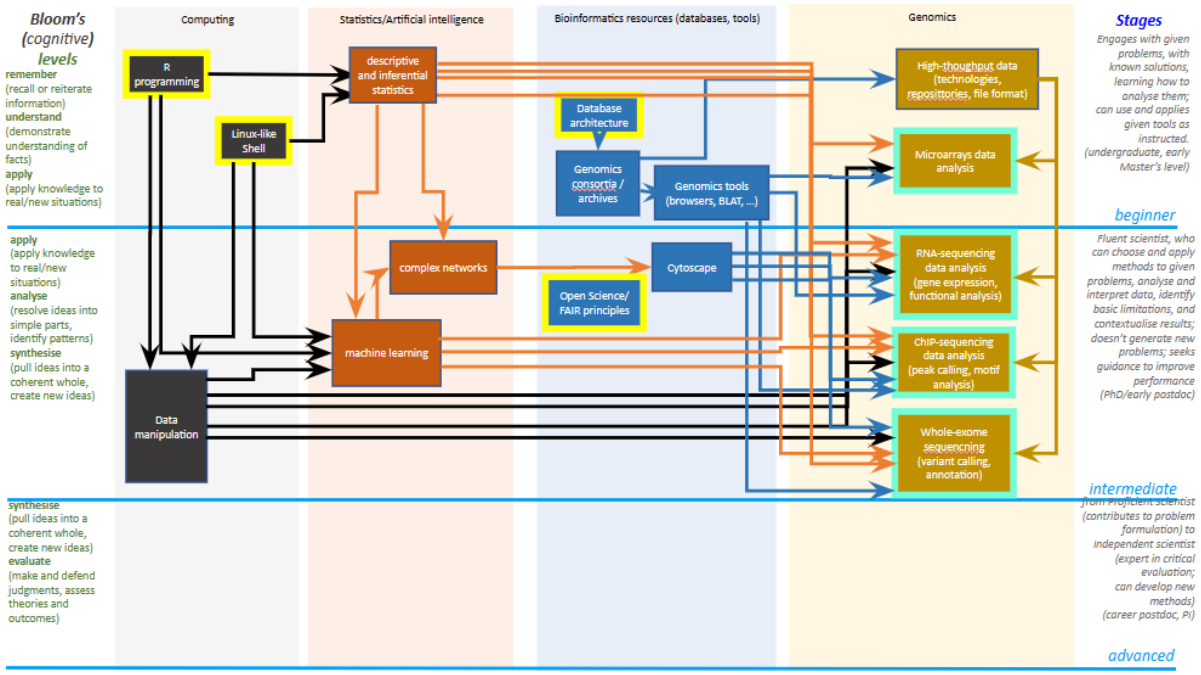


**Figure 5**: A template in progress for example learning paths. With green are highlighted the end-topics of a path. With yellow are indicated topics that have been processed for relevance to the path.[10]

## 4.2 Key examples

### 4.2.1 Data Management / Data Stewardship

HealthRI, in collaboration with the ELIXIR Training Platform community, has designed learning paths for data stewardship, covering the following three roles:

1. Policy oriented data stewardship

   The learning outcome could be: Monitor the compliance to policies and domain protocols by researchers and other stakeholders (evaluating, creating)

2. Research oriented data stewardship

   The learning outcome could be: Stimulate and facilitate the reuse of data by researchers of the organisation (evaluating, creating)

3. Infrastructure oriented data stewardship

   The learning outcome could be: Build a sustainable data-infrastructure and tooling service for the researchers (creating)

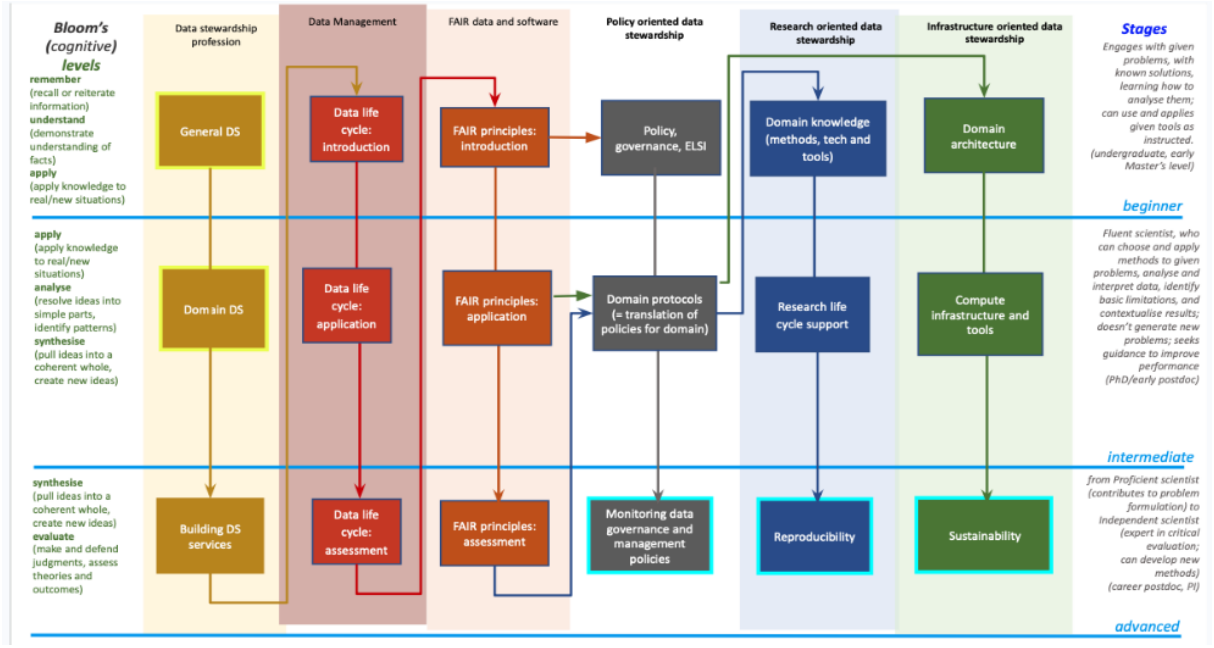


**Figure 6**: Learning paths for data stewards[12].

References for the Data Steward learning path, among others:

- Mijke Jetten, Marjan Grootveld, Annemie Mordant, Mascha Jansen, Margreet Bloemers, Margriet Miedema, & Celia W.G. van Gelder. (2021). Professionalising data stewardship in the Netherlands. Competences, training and education. Dutch roadmap towards national implementation of FAIR data stewardship (1.1). Zenodo. https://doi.org/10.5281/zenodo.4623713
- Salome Scholtens, Mijke Jetten, Jasmin Böhmer, Christine Staiger, Inge Slouwerhof, Marije van der Geest, & Celia W.G. van Gelder. (2022). Final report: Towards FAIR data steward as a profession for the life sciences. Report of a ZonMw funded collaborative approach built on existing expertise (Versie 4). Zenodo. https://doi.org/10.5281/zenodo.7225070
- Research Data Life Cycle; https://ukdataservice.ac.uk/learning-hub/research-data-management/

### 4.2.2 Examples from Commercial Services

Learning paths are becoming increasingly common and accessible to the public beyond the narrow academic context. Usually defined as "*curated selections of training content that guides learners through a more robust development experience*", or *"playlists" that connect content relevant to related skill sets or career paths*"[13], they are being offered through commercial services, such as LinkedIn and Coursera.

[12] Schoots, F., & Jetten, M. (2023, november 8). Learning paths for Data stewards - Presentation for EOSC4Cancer Working Meeting 18 October 2023. Zenodo. https://doi.org/10.5281/zenodo.10084408
[13] https://learning.linkedin.com/resources/career-development/employee-development-learning-paths

Of course, there is still not a standard way of designing or offering these learning paths. For example, there are a lot of differences between the "Become a Data Analyst"[14] learning path in LinkedIn (**Figure 7**) and the Coursera "Data Science Specialization"[15] (**Figure 8**).

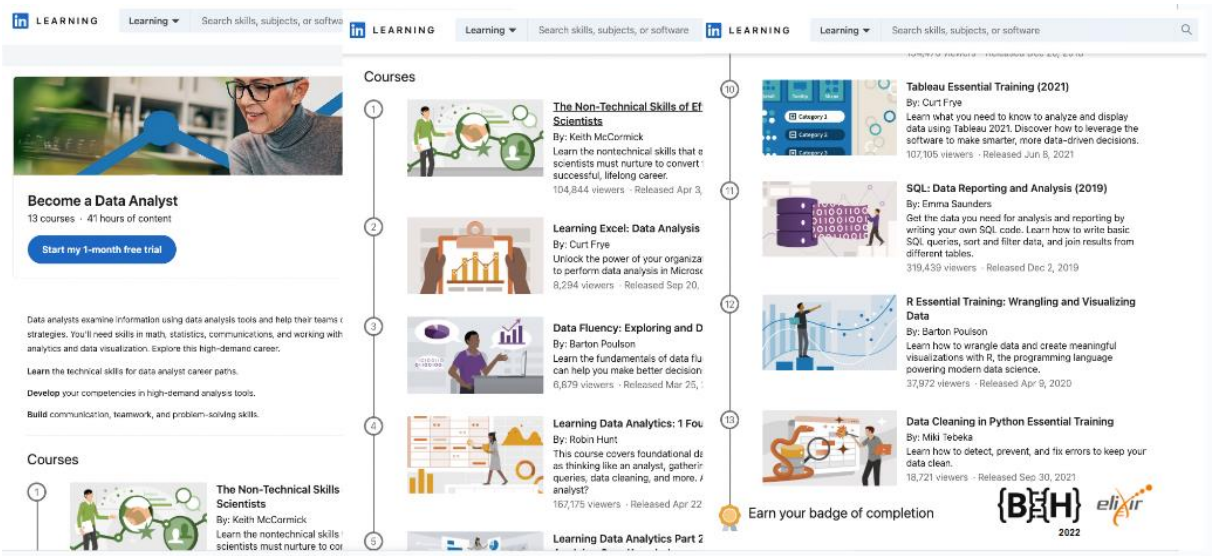


**Figure 7**: The LinkedIn learning path on "Become a Data Analyst"[14]

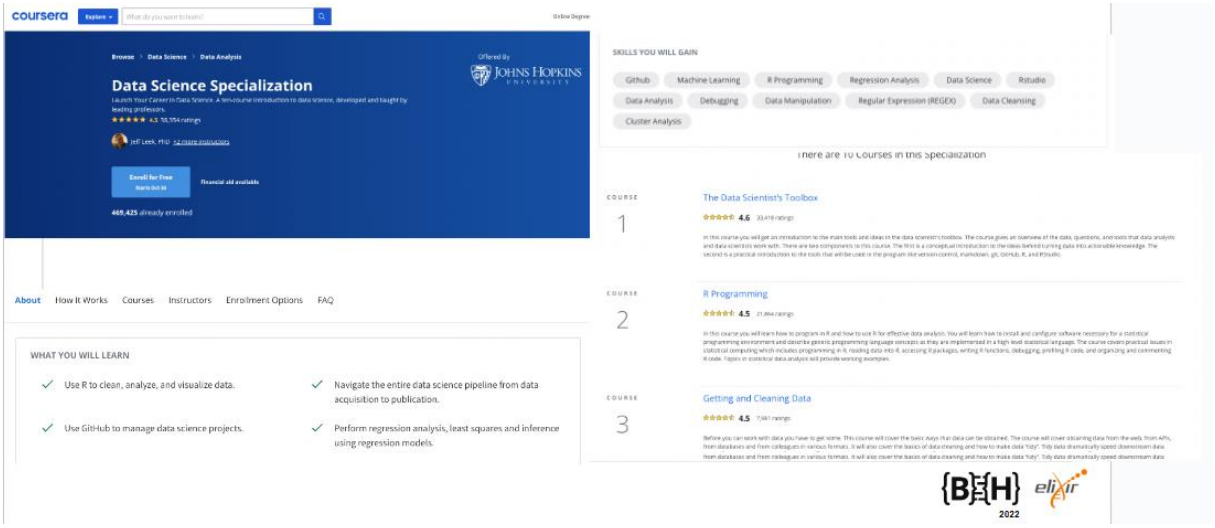


**Figure 8**: The Coursera learning path on "Data Science Specialization"[15]

[14] https://www.linkedin.com/learning/paths/become-a-data-analyst?u=2963594
[15] https://www.coursera.org/specializations/jhu-data-science

# 5 Services supporting Learning Pathways

## 5.1 TeSS

The Training eSupport System (TeSS) is an ELIXIR online training registry[16] that provides a one-stop shop for trainers and trainees to discover online information and content, including training materials, events, and interactive tutorials. TeSS brings together all training events and training material organised by the different organisations and training providers across the life sciences into one easy to use training portal. It also provides opportunities to promote training events and news to contribute to the growing catalogue of materials. TeSS is an integral component of the ELIXIR training platform, which promotes sharing of training materials and exchange of training best practices. Many trainers are passionate about providing open-access training material. Such sharing has several benefits, including providing a record of recognition as an author, offering inspiration to other trainers, enabling researchers to discover training resources for their personal learning path, and improving the training resource landscape using data-driven gap analysis from the bioinformatics community[17].

For trainers, the portal offers an environment for sharing materials and event information and for trainees, it offers a convenient gateway to identify relevant training events and resources. It is possible to manually or automatically register training events and materials on the TeSS for which detailed protocols are provided. TeSS provides protocols[18] for registering, logging in and for searching and filtering content. Following these protocols contributes to easy promotion of training events adding to a growing catalogue of materials.

With improved search-engine optimization and widgets, registering an event on TeSS allows it to appear in more searches increasing its visibility across all websites that display site resources. A key functionality of TeSS is in providing support for the creation, curation and release of Learning Paths - a natural extension of the Collections features of the TeSS. Collections can be thought of as folders in which users may collect related training materials or events, from the full catalogue available within TeSS, to address their specific training needs.

All existing and new cancer data research related training materials created during the course of the project will be connected into dedicated learning pathways, in accordance with the FAIR principles. Currently, a prototype for learning paths is being tested in the TeSS, which will form the basis of the Cancer data related learning paths. Learning Paths on TeSS will be represented as a collection of related training materials (courses/events) collated in a progressive manner, from basic to advanced levels. Individual training materials from different sources will be curated into a logical order to facilitate a progressive and comprehensive learning experience.

[16] Beard, N., Bacall, F., Nenadic, A., Thurston, M., Goble, C. A., Sansone, S.-A., & Attwood, T. K. (2020). TeSS: A platform for discovering life-science training opportunities. Bioinformatics, 36(10), 3290–3291. https://doi.org/10.1093/bioinformatics/btaa047

[17] Garcia, L., Batut, B., Burke, M. L., Kuzak, M., Psomopoulos, F., Arcila, R., … Palagi, P. M. (2020). Ten simple rules for making training materials FAIR. PLoS Computational Biology, 16(5), e1007854. https://doi.org/10.1371/journal.pcbi.1007854

[18] Bacall, F., Apaolaza, A., Andrabi, M., Child, C., Goble, C., Sand, O., & Botzki, A. (2023). Making bioinformatics training events and material more discoverable using TeSS, the ELIXIR training portal. Current Protocols, 3, e682. https://doi.org/10.1002/cpz1.682

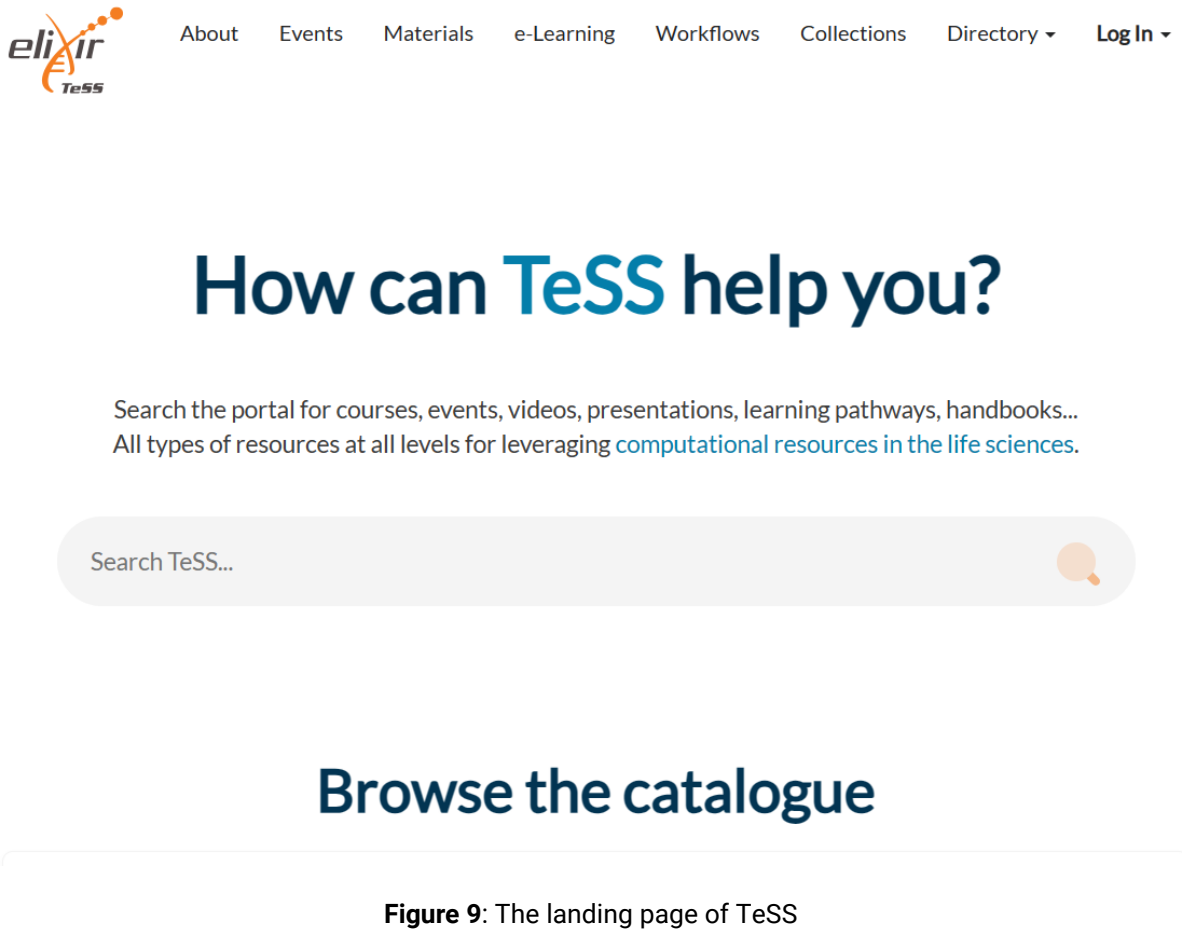


**Figure 9**: The landing page of TeSS

## 5.2 Galaxy Training Network (GTN)

Galaxy is a mature, browser accessible workbench for scientific computing. It enables scientists to share, analyze and visualize their own data, with minimal technical impediments. A thriving global community continues to use, maintain and contribute to the project, with support from multiple national infrastructure providers that enable freely accessible analysis and training services. The Galaxy Training Network (GTN) supports free, self-directed, virtual training with >230 integrated tutorials[19].

As part of the ERASMUS+ Gallantries project[20], GTN organised some of the existing modules into 9 Learning Pathways, in order to provide a guided experience through those modules, based on the background and needs of learners. Each of these learning pathways is designed to function as the program for a 3-5 day training course, and multiple learning pathways may be easily combined into larger curricula e.g. summer schools.

[19] The Galaxy Community, The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update, Nucleic Acids Research, Volume 50, Issue W1, 5 July 2022, Pages W345–W351, https://doi.org/10.1093/nar/gkac247

[20] https://gallantries.github.io

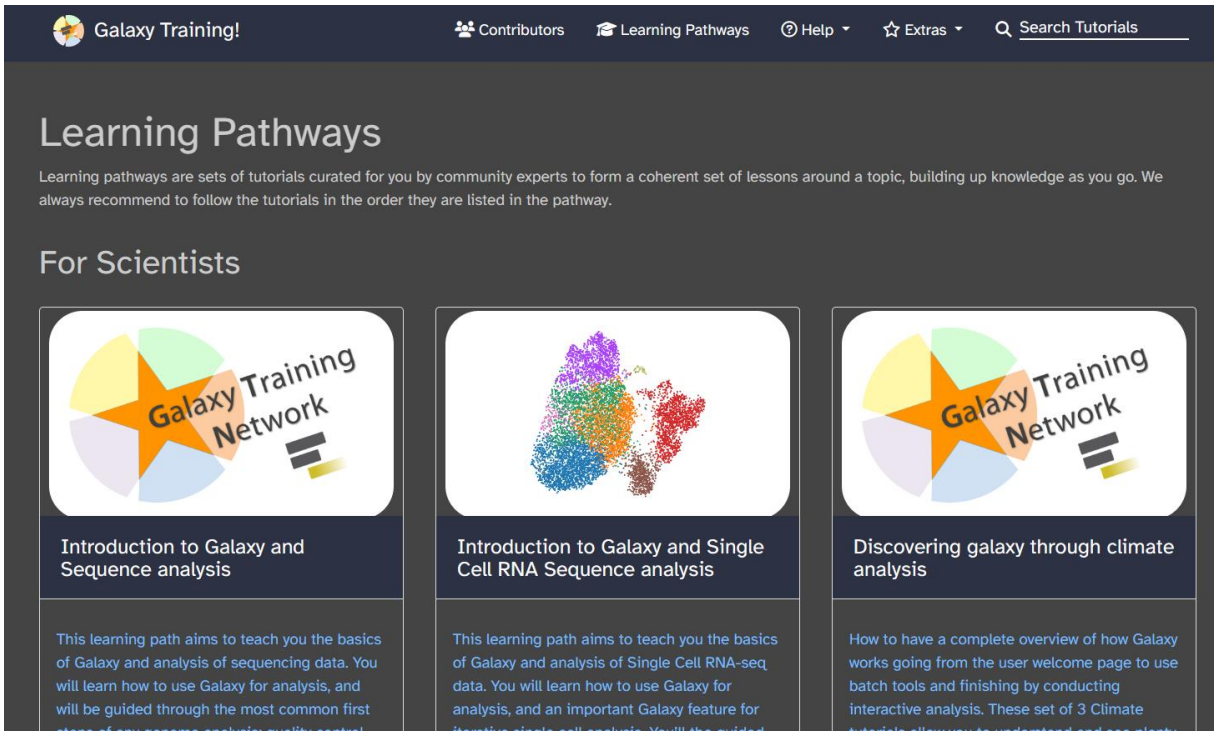


**Figure 10**: The Learning Pathways page on the Galaxy Training Network page[21]

[21] https://training.galaxyproject.org/training-material/learning-pathways/

# 6 Conclusion

Learning Paths are pathways that guide the learner through a set of learning courses or materials to be undertaken progressively to acquire the desired knowledge and skills on a subject. Establishing and implementing Learning Paths will facilitate the learning process in any professional trajectory and career path. Moreover, Learning Paths are meant to continuously evolve and improve, in order to keep pace with the dynamic nature of cancer research data.

This document is a collection of all the information and efforts around the design of Learning Paths. A lot of these activities are still ongoing and will be used as part of the EOSC4Cancer design process as well beyond the date of this document. It is important to implement Learning Paths across different communities and targeting different stakeholders, ultimately raising awareness on the pedagogical aspects and promoting their adoption. To this end, a clear driver of this effort will be the EOSC4Cancer network of cancer research data professionals, that can be established as a community of practice to share knowledge, experiences, and ideas.