Supporting eosc

Skills 4 eosc

Draft Methodology for FAIR-by-Design Training Materials

Lead Partner:	UKIM / GARR
Version:	**1.2**
Status:	**Version open for co-creation - public consultation**
Dissemination Level:	**PUBLIC**
Document Link:	https://doi.org/10.5281/zenodo.7875541

Deliverable Abstract

This document describes a methodology for FAIR-by-design production of learning materials based on the backward instructional process that is extended with additional activities focusing on the implementation of the FAIR guiding principles. A general discussion on important aspect of implementation such as granularity, scope, metadata schema, interoperability and publication in relevant repositories is provided together with a step by step six stage workflow and checklists that help implement the FAIR-by-design process. The outlined methodology will be used as a blueprint for a train-the-trainer course aiming to present the practical FAIR-by-design instructional design.

Co-funded by the European Union

co-funded by UK Research and Innovation

Skills4EOSC has received funding from the European Union's Horizon Europe research and innovation Programme under Grant Agreement No. 101058527 and from UK Research and Innovation (UKRI) under the UK government's Horizon Europe funding guarantee [grant number 10040140]

COPYRIGHT NOTICE

DELIVERY SLIP

Date	***Name***	***Partner/Activity***	***Date***
From:	Sonja Filiposka	UKIM/T2.3	25.04.2023
Moderated by:	Dominique Green	UEDIN/WP2	21.04.2023
Contributor:	Anastas Mishev	UKIM/T2.3	25.04.2023
Contributor:	Vojdan Kjorveziroski	UKIM/T2.3	25.04.2023
Contributor:	Andrea Corleto	GARR/T2.3	21.04.2023
Contributor:	Eleornora Napolitano	GARR/T2.3	21.04.2023
Contributor:	Gabriela Paolini	GARR/T2.3	21.04.2023
Contributor:	Sara di Giorgio	GARR/T2.3	21.04.2023
Contributor:	Joanna Janik	CNRS/T2.3	21.04.2023
Contributor:	Luca Schirru	KUL/ELSI	21.04.2023
Contributor:	Arnaud Gingold	AMU/T2.3	21.04.2023
Contributor:	Christine Chardosek	CNRS/T2.3	21.04.2023
Contributor:	Irakleitos Souyioultzoglou	OPERAS/T2.3	21.04.2023
Contributor:	Carolin Leister	KIT/T2.3	21.04.2023
Approved by:	Emma Lazzeri	GARR	28.04.2023

DOCUMENT LOG

Issue	*Date*	*Comment*	*Author*
v.1.0	***20.02.2023***	***Initial draft prepared***	***UKIM***
v. 1.1	***13.03.2023***	***Internal task review completed***	***T2.3 members***
v. 1.2	***25.04.2023***	***Internal project consultation completed***	***T2.3 members***
…			
v.n			

TERMINOLOGY

https://eosc-portal.eu/glossary

Terminology / Acronym	*Definition*
CC	Creative Commons
COL	Commonwealth of Learning
CSV	Comma Separated Values
DOI	Digital Object Identifier
EOSC	European Open Science Cloud
ETHRD	Education and Training on Handling of Research Data
EU	European Union
FAIR	Findability, Accessibility, Interoperability, Reusability
FLOSS	Free/Libre/Open Source Software
FOSS	Free and Open Source Software
GNU	General Public licence
GOBLET	Global Organisation for Bioinformatics Learning, Education and Training
H5P	HTML5 Package

HTML	HyperText Markup Language
ID	Identifier
IEEE	Institute of Electrical and Electronics Engineers
IG	Interest Group
IPR	Intellectual Property Rights
JSON	JavaScript Object Notation
JSON-LD	JavaScript Object Notation for Linked Data
LMS	Learning Management System
LOM	Learning Object Metadata
MOM	Mason OER Metafinder
MOOC	Massive Open Online Course
ND	No Derivative
OER	Open Educational Resources
OS	Open Science
PDF	Portable Document Format
QA	Quality Assurance
RDA	Research Data Alliance
RDF	Resource Description Framework
SA	Share Alike
SCORM	Shareable Content Object Reference Model
SGDR	Sui Generis Database Right
SSHOC	Social Sciences & Humanities Open Cloud
UNESCO	United Nations Educational, Scientific and Cultural Organization
URL	Uniform Resource Locator
US	United States
USG	University System of Georgia
WG	Working Group

WIPO	World Intellectual Property Organization
XML	eXtensible Markup Language
YAML	YAML Ain't Markup Language
Aggregation	Hierarchical level of composition of multiple learning objects.
Application profile	An application profile is a document or set of documents that contains functional requirements, domain model, description set profile, and syntax guidelines and data formats.
Attribution	Acknowledgement as credit to the copyright holder or author of a work.
Authentication	The process or action of verifying the identity of a user.
Authorisation	The process of giving someone permission to do or have something.
Backward instructional design	Begins with the learning objectives and then proceeds ”backward” to create content that achieves those desired goals.
Citation	A reference to a published or unpublished source of information.
Continuous improvement	The ongoing improvement of learning materials through incremental and breakthrough improvements.
Controlled vocabulary	An organized arrangement of words and phrases used to index content and/or to retrieve content through browsing or searching.
Copyright	Type of intellectual property that protects original works of authorship as soon as an author fixes the work in a tangible form of expression.
Courseware	Online resource that students can use to learn and study including learning materials, activities, quizzes, collaboration tools, etc.

Final output format	Final representation of the learning object ready for consumption by learners, most often no longer manually editable.
Git	A distributed version control system capable of tracing changes in computer files.
GNU	A series of free software licenses that guarantee end users the four freedoms to run, study, share, and modify the software.
Granularity	Refers to the "size" or "extent" of a learning object.
Human-readable format	Any encoding of data or information that can be naturally read by humans.
Instructional design	The creation of learning experiences and materials resulting in the acquisition and application of knowledge and skills.
Instructor / Instructional designer	A learning expert who can use their knowledge of the principles of learning and instruction to find the optimal method of instruction.
Instructor kit/Facilitator kit	Accompanying material that aims to help facilitate the process of other instructors reusing the learning material.
Intellectual Property Rights	The exclusive rights given to persons over the creations of their minds. According to the World Intellectual Property Organization [R1] 'IP is often divided into two main categories: Industrial property includes patents for inventions, industrial designs, trademarks and geographical indications. Copyright and related rights cover literary, artistic and scientific works, including performances and broadcasts.'
Intermediary file format / source format	The format which is used during the development phase of the learning objects.
Internal quality assurance	The processes and procedures within institutions to review, evaluate, assess or otherwise check, examine or ensure the quality of the learning provided.

Learner	A person (ex. student, trainee) who is learning a subject or acquiring a skill.
Learning content	The topics, themes, beliefs, behaviours, concepts and facts, often grouped within each subject or learning area under knowledge, skills, values and attitudes, that are expected to be learned and form the basis of teaching and learning.
Learning context	A short summary or concept description that defines the learners' perception of the material and the requirements.
Learning object	Any digital resource that supports learning developed around a single learning objective defined as a package of a lesson, activity and assessment with a concrete learning outcome. This is the minimum resource on which the FAIR principles are applied.
Learning objective	A statement that clearly defines the expected outcome as a result of the learning activities and assessments.
Learning path	The chosen route taken by a learner through a range of learning activities, which allows them to build knowledge progressively.
Learning platform	A type of software that enables instructors to create and deliver courses online.
Learning / Training / Instruction materials / resources	Any types of materials that are used to support and enhance, directly or indirectly, learning and teaching with their main goal being to help achieve the desired learning objectives. The full scope of learning materials includes learning content, tools and implementation resources.
Licence	The (exclusive or non-exclusive) licence of use of IP rights from the owner to a third-party, short of an assignment of all rights.
Licence compatibility	A legal framework that allows for content with different licenses to be distributed together. Usually

	used for software, but also applicable to content licensed under CC or related licenses.
Lifelong learning	The practice of continuing to learn throughout one's entire life, especially outside of or after the completion of formal schooling.
Link-rot	Problem of no longer accessible hyperlink due to the resource being moved to a new location, deleted, or permanently made inaccessible for any other reason.
Machine-readable format	Structured data in a format that can be processed by a computer. Most popular formats include JSON, XML, YAML
Metadata longevity plan	Policy and procedures for digital archiving, backup schedules, and preservation of fair objects and their metadata.
Metadata schema	Metadata schema outline the overall structure for the metadata. It is a logical plan showing the relationships between metadata elements, normally through establishing rules for the use and management of metadata.
Ontology	A set of concepts and categories in a subject area or domain that shows their properties and the relations between them.
Open Educational Resources	Learning, teaching and research materials in any format and medium that reside in the public domain or are under copyright that have been released under an open license, that permit no-cost access, re-use, re-purpose, adaptation and redistribution by others.
Open file format	A file format for storing digital data, defined by an openly published specification usually maintained by a standards organization.
PID	A long-lasting reference to a document, file, web page, or other object.

Prerequisite	Anything the learner needs to know or understand first before learning or understanding the offered content.
Syllabus / Content Concept Map	A document that presents the purpose of the learning content with precise description of what is to be learnt, how and when under the assumption of the defined prerequisites.
Trainer	A learning expert who delivers training.
Training catalogue	Provides a description of the training services and materials offered along with the related policies and procedures in regard to such training.
Training toolkit	A tool for trainers and more generally educators to enhance their competence in providing training.
Versioning	The creation and management of multiple learning materials published releases, all of which have the same general function, but are improved, upgraded or customized.

Contents

Executive summary

Training is a vital part of the European Open Science Cloud (EOSC) system aiming to upskill a large community of various stakeholders with the competences necessary to both build and use data and services. It is therefore important that the EOSC training community is able to efficiently collaborate ensuring the delivery of high-quality up-to-date trainings and learning opportunities. To help achieve this goal, Task 3 of Work Package 2 within the Skills4EOSC project has developed a methodology for FAIR-by-design learning materials that will ensure maximum reusability of developed learning materials within the community, and with that higher-quality materials.

The proposed methodology builds on the previous work done in other EOSC projects, while incorporating best practices and lessons learnt from related activities such as implementation of learning platforms, development of self-paced courses, definitions of metadata schemas for training materials and integration of training catalogues. All necessary steps to ensure the production of FAIR-by-design learning materials are outlined in a six-stage workflow that extends the traditional instructional design process with additional activities aiming to incorporate within the FAIR principles.

Each stage of the workflow discusses the relevant aspects of learning material development blending learning models, materials and methods with the FAIR requirements. In this way an efficient, lean approach is proposed for instructors that are guided through each step of the design process helping them expand their instructional design skillset with FAIR relevant competencies. The approach empowers instructors to efficiently handle the following: legal issues, such as intellectual property, licensing and attribution; material description and referencing, by choosing and using a metadata schema with relevant controlled vocabularies together with the use of persistent identifiers offered by different types of possible repositories for the learning materials; interoperability issues, by learning how to combine different file formats and tools and understand the difference between final and intermediate content packages; design and structuring of the learning

content, thus ensuring the most appropriate level of granularity for maximum reusability; and accessibility issues, by not just defining access rights, but also ensuring usability by people with disabilities.

The workflow stages also include a number of quality assurance checkpoints together with the activities on defining and handling feedback from different stakeholders supporting the process of co-creation of the learning materials with other peers. The complete process encompasses the principles for continuous improvements where the gathered feedback can be fed back to the process helping improve the quality of the produced learning material.

1 Introduction

Training and skills development are the cornerstone for building an effective ecosystem wherein the users and providers can take full advantage of the new possibilities offered by the European Open Science Cloud (EOSC). Through the federation of data and research infrastructures across Europe, EOSC aims to provide seamless access and research data reuse in a robust, secure, scalable and flexible way [R2]. To achieve this vision, in parallel with the advancements of technologies and development of resources and services, it is imperative that all traditional activities that are part of the research lifecycle are transformed using Open Science (OS) approaches. This entails empowering a diverse set of stakeholders with skills that can help them understand and employ the benefits of OS and EOSC services. And, as the EOSC ecosystem is continuously evolving, the need for continuous upskilling will remain. Having this is mind, the training and skill development activities should be implemented in a sustainable manner by building and supporting growing communities of trainers and developing high-quality reusable training materials.

The Findability, Accessibility, Interoperability and Reusability (FAIR) guiding principles [R3] are at the heart of the EOSC activities defining the features that all data resources, tools, services and infrastructures should have to promote discovery and reuse by third-parties. The principles act as a guide when choosing specific implementation choices aiming to ensure that the created digital artefacts are Findable, Accessible, Interoperable, and Reusable. The approach to upskilling stakeholders to engage in EOSC should follow the same values, in particular when it comes to the development of materials that can be used for training sessions, or self-paced learning about different aspects of OS and EOSC. The benefits of focusing on the development of FAIR learning materials go beyond the long-term investment for the EOSC training community including other aspects such as:

- Expanded base of learners
    - Encompassing not just targeted trainees or OS students, but any interested party that would like to use the provided learning

material, thus not only supporting, but actively boosting lifelong learning experiences;

- Improved learning process
    - As FAIR learning materials mean that learners can easily find and access learning content, obtain more in-depth understanding of an offered course or training before actual enrolment, or go back to refresh their knowledge on a given topic;
- High-quality learning materials
    - FAIR learning resources can be adapted and revised, and in this way more easily kept up to date, translated and localized to a specific context;
    - The metadata that accompanies the learning materials offers clear information regarding licensing fostering reuse through adaptation and development of enhanced learning content;
    - Existing learning resources can be revised and reused to build various learning aggregations such as learning paths or certification requirements;
- Sustained network of instructors
    - Trainers and teachers can create or review learning materials in a collaborative fashion.

Training/learning materials development regarding OS and EOSC are still in its early stages with many initiatives and parallel activities needing to be aligned and harmonised. This process has already started with the activities of the Training and Skills EOSC Working Group [R4] and continues with the activities in the new EOSC projects such as Skills4EOSC as well as the new task forces such as the Upskilling Countries to Engage in EOSC Task Force [R5]. On the journey to achieving a FAIR-by-design approach to the development of learning materials for EOSC, there are many challenges than need to be tackled:

- Finding existing FAIR learning materials on a given topic
    - In the past years there has been significant investment in training on the topics of OS and EOSC, and these efforts have produced an initial body of available learning materials. However, these are still

scattered on different platforms and repositories, and providing a single point for searching and accessing is still an open issue that is currently being tackled by projects such as EOSC Future [R6];

- Learning materials formats
    - Most of the available learning materials can be found in closed formats making them difficult to be reused by other instructors, especially when adaptation is needed. Also, the available content is mostly slide handouts, video recordings of webinars or short packaged courses, while the accompanying material such as exercises, quizzes, instructional guides, etc. are difficult to find or extract;
- Extra effort
    - It is clear that making learning materials FAIR adds a considerable overhead on the already lengthy process of development of new learning resources. Thus, awareness is needed that additional time and competencies are required when aiming to produce high-quality FAIR learning materials with enough granularity to ensure maximum re-usability.

Aiming to further support the EOSC training and skills community, one of the goals of the Skills4EOSC project is to develop a comprehensive FAIR-by-design methodology that will cover all aspects of the process of developing FAIR learning materials. This methodology is envisioned as a tool that can be used by the community of training/learning materials designers helping them upskill their traditional design process by paying attention to ensure FAIRness of the produced learning content.

The purpose of this document is to introduce the concept and principles of the FAIR-by-design methodology, including all relevant aspects and necessary skills. It discusses the concept of modelling and describing learning materials, the related work regarding training metadata and materials development, to then propose a FAIR-by-design methodology for developing training materials by augmenting the instructional design process. Throughout the lifetime of the Skills4EOSC project, the feedback gathered from relevant internal and external stakeholders as well as training sessions on how to

practically implement the proposed methodology will be used to further improve the methodology and its value to the EOSC training community.

2 Importance of FAIR-by-design approach

Currently, implementing the FAIR principles for learning materials is mostly taken as an afterthought, usually triggered by an issue related to one or more of the FAIR principles. Turning digital learning materials into FAIR learning materials after their development can be very tedious, time-consuming and error-prone because even if the original content author is doing the work, one must go through the whole material once more and gather additional information such as attribution and licensing or reused learning objects. If the original editable materials have not been versioned and stored properly, it will be very difficult to find the newest version to add and/or extract the necessary information. Then the new final learning package needs to be regenerated and published on a learning platform or other location where it will be available for consumption. Additionally, there will be a number of other steps involved such as defining the necessary metadata, adding facilitator package, cataloguing, etc. In essence, the FAIR-ification process will require recreating a whole new version of the learning materials.

Thus, it is essential that the FAIR principles are incorporated within the design process. In this way, unnecessary duplication of work can be avoided, and all necessary considerations have been properly addressed. That means that the FAIR-ification is being implemented throughout all stages of the learning materials development process from the inception to the release for use.

2.1 Previous work

When it comes to implementing the FAIR principles to research datasets, there have been many projects and initiatives such as FAIRsFAIR and GO FAIR that were tackling different aspects and challenges of accomplishing the goal of FAIR research data production. It is, however, not so straightforward to apply the FAIR principles to other research output as it is recommended for datasets. Lately, there have been substantial efforts on extending the FAIRness idea to other types of research output as highlighted by the work done by the Research Data Alliance that has extended the application of FAIR principles to research software [R7].

The use and development of learning materials has very specific features that needs to be adequately addressed in order to truly implement the FAIR principles, particularly when it comes to describing the materials using a metadata schema or combining materials with dissimilar licensing information. This is why the specific issue of FAIR learning materials has been of importance in the work of different EOSC projects, groups and alliances.

For example, the ENVRI-FAIR project has aimed at developing FAIR training materials integrated into a training platform and catalogue and its deliverable outlines some example steps taken to produce the materials [R8]. Substantial efforts towards the creation of high-quality training materials have been done by ELIXIR in collaboration with Global Organisation for Bioinformatics Learning, Education and Training (GOBLET) providing a training platform, but also a comprehensive training toolkit and other important training related information targeting specifically the bioinformatics community [R9]. They have also worked on a metadata standard for describing training metadata in bioinformatics. The FAIRsFAIR project has published a report on harmonising metadata for FAIR training materials exchange [R10], and this work has later been picked up and continued by the Research Data Alliance. EOSC Synergy is another example project that has produced an online training handbook [R11] that serves as a facilitator kit for online training development. At the moment one of the most important efforts on the topic of making FAIR training materials is presented in the "Ten simple rules for making training materials FAIR [R12] that lays out the essential requirements needed to ensure that the training materials are FAIR. The FAIR-by-design methodology proposed in this document aims to build upon all of the work done already in the EOSC community and help the challenges of practical implementation of the FAIR principles in the EOSC training community.

As the FAIR principles do not require the data in question to be Open, there is a growing community, particularly in the US, that goes beyond the FAIR requirements and fosters the development of Open Educational Resources (OER) [R13] which are supported by UNESCO. In essence, OER are learning materials that implement all of the FAIR principles using their own specific

metadata schema for description, and, in addition, are required to be fully open for use and reuse. Some of the European projects, such as TRIPLE [R14], have adopted the OER approach and aimed towards fully open materials. All recommendations and available guidebooks on how to make OER have been taken as valuable input in the definition of the proposed FAIR-by-design methodology, with the option on having the finally produced materials fully open for use and reuse as a recommended practice. Thus, using the proposed methodology one can choose the level of openness of the materials with which they are comfortable with making the methodology applicable in various scenarios by supporting the idea of as open as possible, as closed as necessary.

3 Learning materials description and modelling

To ensure a consistent, yet flexible, approach to the development of a FAIR-by-design methodology for EOSC related learning materials, a well-defined scope and modelling of learning materials are needed. This chapter is a formal introduction to the theories and frameworks that one can use to apply the FAIR principles to new or pre-existing EOSC learning materials.

Herein, we adopt the definition of learning materials or learning resources as any types of materials that are used to support and enhance, directly or indirectly, learning and teaching [R15] with their main goal being to help achieve the desired learning objectives. Note that in the general literature learning resources are also known as teaching-learning material, instructional materials or teaching aids. It has been shown that the utilisation of a large assortment of learning materials improves comprehension and improves the ability to learn in both group or independent setting, on premises or online [R16]. Hence, today there are many different types of learning materials including:

- printed materials such as handouts and manuals;
- audio-visual materials such as slides, images, videos, podcasts, and other multimedia;
- interactive materials in the form of learning applications on computers, tablets or smartphones;

all of which are used in the instructional activities, including active learning and assessment. In line with the EOSC vision and the overarching digital transformation, the main focus of the FAIR-by-design methodology is on the digital learning materials. Digital materials are essential not only in the e-learning environment, but also in the more traditional learning settings such as classrooms, face to face trainings and workshops; all of which are seen as a potential learning venue within the EOSC ecosystem.

In addition, in line with [R17], the full scope of learning materials is considered to include:

- **Learning content**, such as courses, modules, learning units, etc. This includes all types of courseware such as learning materials, presentations, guides, case studies, activities, quizzes, etc.
- **Tools** that represent the software necessary to develop, use and deliver the learning content such as learning management systems, content development tools, online learning communities, etc.
- **Implementation resources** that are needed to support the development and promotion of the learning content such as best practices guides, licensing and copyright used for promotion and reuse, publishing standards, etc.

This entails that each learning resource needs to be accompanied together with the corresponding tools for its design and consumption as well as well-defined conditions for its use as presented in Fig. 1.

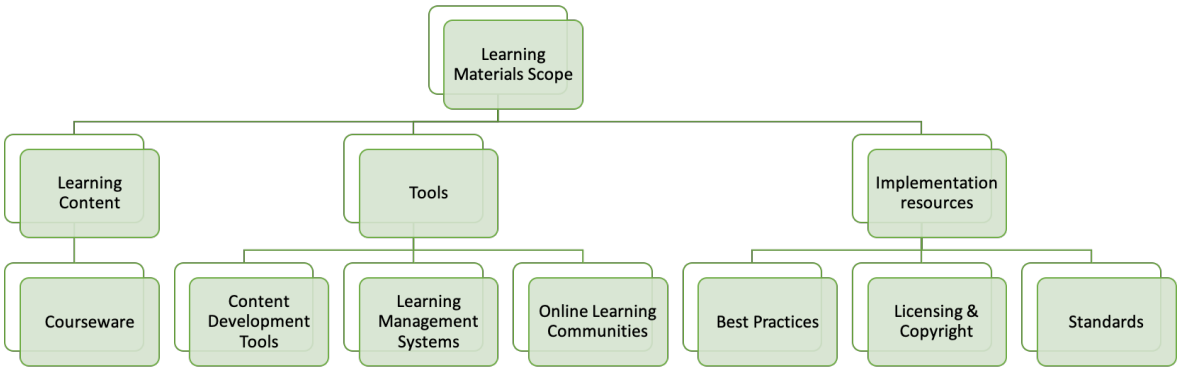

Fig. 1 - Learning materials scope

One of the main decisions that need to be made when ensuring FAIRness of learning materials is to define the level of granularity on which the FAIR principles will start to be applied for a given set of learning materials. In other words, the main question is what is the minimum sized package of digital learning materials that is to be subjected to the FAIR principles? Is it a course (or training), or a unit within a course (session within the training), or a single digital resource? If the granularity is too high, then there is a large overhead of metadata provisioning and cataloguing for a vast amount of FAIR learning resources with very little context related to them. On the other hand, low granularity will significantly reduce the effort of cataloguing, but also the

possibility to combine different learning resources in order to create new learning content out of existing learning materials due to the inability to adopt only selected parts of a given material. To ensure flexibility and minimise the overhead, the minimum viable package of learning materials on which one can apply the FAIR-by-design methodology is based on the instructional design concept of learning object [R18].

Thus, in the most general sense, we define a FAIR learning object as any FAIR digital resource that supports learning developed around a single learning objective. At the very core of this approach is the idea that the learning object, accompanied with suitable descriptive metadata, can be used as a common building block for the development of more complex learning content by reusing it, re-purposing it, and potentially revising it.

Furthermore, to facilitate the reuse of learning objects, the best practice is to define a learning object as a package of a lesson, activity and assessment with a concrete learning outcome. Note that a similar approach is adopted in the OER community [R19]. Other communities put even more heavy requirements on the minimum content of a learning object. Following the best practices presented in [R20], a learning object should include:

- Meaningful title (and subtitle);
- Single, specific learning objective;
- Target audience;
- Time required to complete the content;
- Guide for instructors and learners;
- Learning content;
- Self-evaluation;
- Final test to be used for formal evaluation.

Using this approach, the process of combining FAIR learning objects into higher level FAIR learning resources can be conceptualised in a fashion similar to the well-known Learnativity aggregation model (Fig. 2). The most granular element in the system is a raw data element that represents a single digital file with no context. Once the learning context such as a summary or concept description is provided for this file, it transforms into an information

object. One or more information objects that provide the learning content around one learning objective become a learning object. These objects can then be combined into different aggregates or assemblies to define lessons, or units, which are in turn combined into collections that can represent courses or learning paths. By defining the learning object as the minimum resource to which we apply the FAIR principles, one avoids the necessity to make raw elements and information objects FAIR, while enabling the creation of FAIR aggregates and collections of various sizes and complexities.

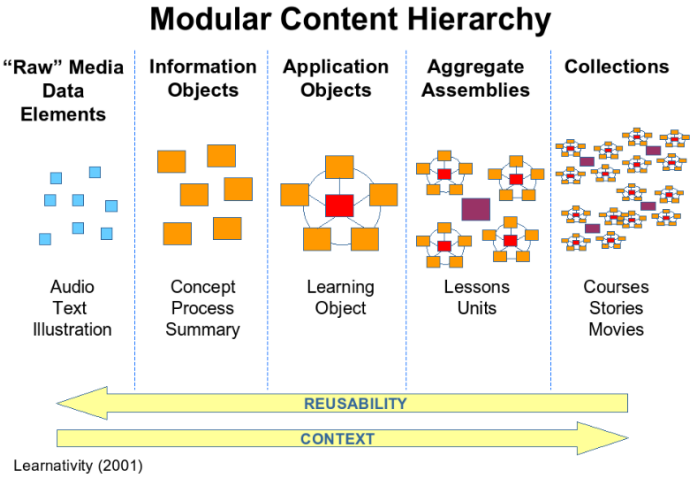

Fig. 2 - Learnativity aggregation model [R21]

In this sense, learning objects aim to facilitate FAIRness of learning materials by enabling the design of small self-contained units of learning content that can be then reused in different courses and other learning materials or even large programmes. Following these concepts, FAIR learning objects exhibit the following characteristics:

- **Findable** - the learning object is the lowest hierarchical level of findability of learning materials in the EOSC ecosystem and is thus the lowest hierarchical level that can be described with metadata and catalogued;
- **Accessible** - the full scope (content, tools and implementation resources) of the learning object should be accessible to both learning producers and consumers in the EOSC ecosystem;

- **Interoperable** - with a well-chosen scope (content, tools and implementation resources), the learning object can be consumed on multiple platforms;
- **Re-usable** - each learning object can be put in a wider context based on the specific learning requirements of a particular aggregate course, unit or module in the EOSC ecosystem.

Note that the definition of the aggregation model is such that these characteristics are also applicable to any higher-level aggregation of FAIR learning objects.

This approach to composability of learning objects is well aligned with other efforts such as the IEEE LOM [R22] wherein there are four levels of granularity:

- L1 - a unit of learning, atomic material, indivisible learning material such as an image;
- L2 - a collection of L1 objects, i.e. lesson, such as a web page that is a set of an HTML file and a number of images;
- L3 - a collection of L2 objects, i.e. course, such as a web site with a number of pages;
- L4 - highest level of granularity, such as study programme or learning path obtained as a collection of L3 and/or L4 objects.

The FAIR learning object in this case corresponds to the L2 granularity.

Similarly, the Shareable Content Object Reference Model (SCORM) [R23] that represents a collection of standards that aim to promote interoperable, accessible, and re-usable learning content defines a content model that consists of:

- Assets - smallest piece of learning content;
- Shareable Content Objects (SCOs) - aggregation of assets that communicates with a virtual learning environment;
- Content aggregation - structured map of learning resources.

The SCORM content aggregation is composed of assets, SCOs, definition of their order and metadata that describes the entire aggregation and its

individual components. Thus, in the proposed model, a SCORM content aggregation can be considered to correspond to a FAIR learning object.

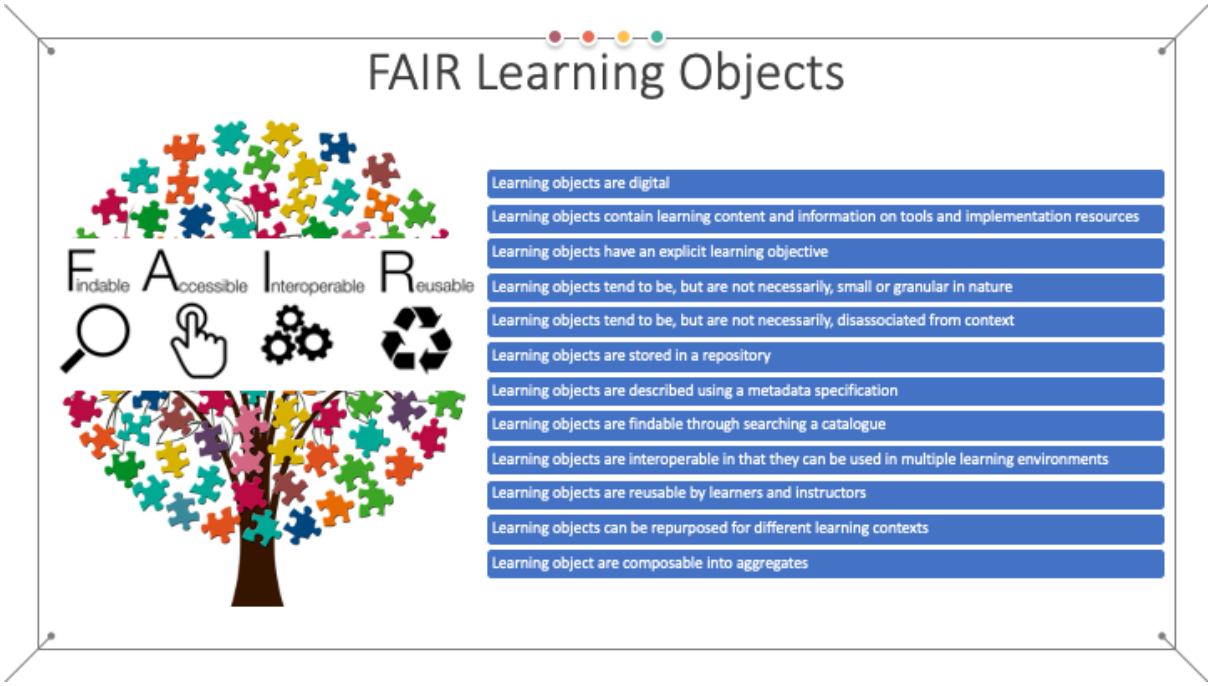

Fig. 3 - FAIR learning objects characteristics

In summary, the FAIR-by-design methodology strives to ensure the implementation of the following specific characteristics of FAIR learning objects (Fig. 3):

- Learning objects are digital;
- Learning objects contain learning content and information on tools and implementation resources;
- Learning objects have an explicit learning objective;
- Learning objects tend to be, but are not necessarily, small or granular in nature;
- Learning objects tend to be, but are not necessarily, disassociated from context;
- Learning objects are stored in a repository;
- Learning objects are described using a metadata specification;
- Learning objects are findable through searching a catalogue;

- Learning objects are interoperable in that they can be used in multiple learning environments;
- Learning objects are reusable by both other instructors and learners;
- Learning objects can be repurposed for different learning contexts;
- Learning object are composable into aggregates.

3.1 Describing learning materials

3.1.1 The importance of metadata

Metadata plays a central role in the implementation of the FAIR principles and enables easier sharing of data and material. By associating relevant metadata information to each learning resource, characteristics relevant to all 4 FAIR aspects can be identified at a glance. Descriptive metadata that is understandable by humans and readable by machines aids the findability of the content through search engines or specialized catalogues. Including information regarding the access rules and associated licence improves the accessibility of the resources. Adding a summary description to each item and modelling its relation to other items through the use of metadata illustrates the interoperability of the given learning material, thus boosting its reusability value.

3.1.2 Metadata schemas

For metadata to be effective and to fulfil the described objectives, its structure must be consistent and unambiguous, as well as adhere to a widely used schema. According to ISO, metadata schema [R24] is "a logical plan showing the relationships between metadata elements, normally through establishing rules for the use and management of metadata specifically as regards the semantics, the syntax and the optionality (obligation level) of values". At present a number of metadata schemas focused on learning resources exist, with varying verbosity, tackling different aspects and subject areas. Schema.org [R25] is one of the most popular examples and its versatility allows the same vocabulary to be used for different types of resources. When describing learning resources, the relevant type is "CreativeWork" [R26]. Being a community led effort, it is also possible to

extend existing types or derive new ones. Both "Course" [R27] and "LearningResource" [R28] (the latter of which is still not fully integrated) share the same "CreativeWork" parent.

Bioschemas [R29] along with the Open Educational Resources Schema (OER) [R30] are two additional examples which are based on the work of Schema.org. Both extend the existing vocabulary with additional terms and declare new types. Bioschemas is primarily focused on describing datasets, software and training materials related to life sciences. OER is aimed at traditional learning materials, introducing granular types, such as: "Assessment", "Course", "Quiz", "Project", etc. The available context specific terms such as "termOffered", "department", "program", underscore the main applicability of this schema as one for formal education institutions. The European Life Sciences Infrastructure for Biological Information (ELIXIR) [R12] has also defined a metadata set to aid the trainees of their training platform to better identify the resources relevant to them. It is less verbose than the ones already described, and includes 13 core fields of information, describing general information, prerequisites, and outcomes for each resource.

Whenever discussing existing schemas or defining a completely new one, the existence of a significant trade-off needs to be recognized. On one hand the addition of new fields aids the overall descriptiveness and might increase the findability and reusability value of the described resources. However, on the other hand, mandating the presence of a large number of distinct fields hinders the adoption of the given schema, making it more difficult to ensure conformity of existing or new material. To solve this problem, the Education and Training on Handling Research Data Interest Group Research Data Alliance (ETHRD-IG RDA) task force developed the minimal RDA metadata set [R31]. This metadata set has been derived through the analysis of six existing metadata schemas, some of which were described above, with the end goal of creating an easily adoptable set of metadata elements. It is expected that resource creators would benefit from such a metadata set, allowing them to describe their learning resources when making them publicly available. The RDA metadata schema consists of 14 different fields, divided into 3 different

categories of information: descriptive, access, and educational. It is the recommended metadata set to be used both for existing and new learning materials. An even more restricted profile aimed at establishing faster conformity for existing materials consisting of a subset of only 3 mandatory fields has also been proposed [R32]. These fields are:

- Title - a human readable name of the learning resource
- Author(s) - the name of the entities authoring the learning resource
- URL to resource - a URL resolving either to the learning resource itself or to a dedicated page which includes additional contextual information including a direct link to the underlying resource.

A number of training platforms are currently actively evaluating the RDA minimal metadata set, such as OpenPLATO [R33], the training catalogue of the SSHOC project [R34], and the NI4OS Training Platform [R35]. As the current EOSC efforts on defining a common metadata schema for learning resources are adopting the proposed minimal schema, the use of the RDA minimum metadata schema for learning resources [R31] is recommended taking into account that close attention should be given to its future development.

3.1.3 Proposed extensions to the recommended RDA metadata schema

Agreeing on an existing, well-defined, and descriptive metadata set is essential for reusability of materials, and their findability through general purpose search engines and specialized catalogues such as the currently being developed by EOSC Future [R36] which is envisioned to become the overarching training catalogue for the EOSC community. To aid the existing effort, we recommend the extension of the RDA metadata set with two additional fields, "isPartOf" and "isBasedOn". Both of these fields are already part of Schema.org and its derivatives. The allowed value for the fields in this case would be a URL to the respective resource. The inclusion of these two fields would increase the number of minimal metadata elements to 16, but with the added benefit of being able to better model hierarchical relationships between learning materials, and the findability of related

content. In essence, this will allow the metadata information to better reflect the relationships between learning objects and allow traceability of information back to its source.

Additionally, the set of values possible for the licence field could be further restricted, mandating that it only be a URL to the text of the associated licence. This would make it easier for machines to understand the field and avoid ambiguities which might arise as a result of inconsistent spelling or omitting a version of a given licence.

3.1.4 Controlled vocabularies as a framework for metadata values

To ensure the descriptive value of the additional information associated with each learning resource, and to make it consistent across different applications, metadata schemas restrict the values that a given field may contain. This can either be in the form of mandating its type - e.g., differentiating between a text or a number field; its cardinality, describing how many times it can be repeated; or its content altogether, specifying a set of pre-approved values from which the author or administrator can choose (controlled vocabulary). It is important to recognize that even though this behaviour might seem restrictive at first, it is necessary to ensure interoperability and in-ambiguity between platforms using the same metadata schema, while also providing uniform experience to the users. Content creators are strongly encouraged to adhere to the outlined guidelines. The document describing the recommended RDA metadata schema also includes information related to such restrictions [R37]. At present the following fields of the RDA minimal metadata schema have controlled vocabularies: Primary Language, Version Date, Resource URL Type, Target Group, Learning Resource Type, Access Cost, Expertise Level.

3.1.5 Metadata representation formats

Metadata can be even more relevant for machines than it is for humans. By ensuring that metadata for learning materials is provided in a machine-readable format, it can be ensured that it will be interpreted in the desired

context by automated tools such as search engines, crawlers, link generators, and bots. To achieve this, learning infrastructures should be capable of serving the metadata information in a variety of formats, such as: unstructured, Comma-Separated Values (CSV), JavaScript Object Notation (JSON), Extensible Markup Language (XML), YAML Ain't Markup Language (YAML). The unstructured representation is most relevant for humans and can be provided in a visually appealing way, disregarding readability by machines. The CSV format can be beneficial for doing bulk information dumps, due to its simplicity, easy understandability by humans, and interoperability with existing software. However, the formats most popular today for data interchange between machines are JSON, XML and YAML. Most training catalogues and learning resource aggregators today make use of at least one of these three formats, in order to keep the metadata information in sync across the various training portals.

A number of metadata schemas have also introduced application profiles using the main standards for linked data including JavaScript Object Notation for Linked Data (JSON-LD) and Resource Description Framework (RDF). Such application profiles allow machines not only to read the data, but also interpret it and understand the context in which it is provided. An application profile is currently in the development phase for the RDA schema [R38].

3.1.6 Metadata longevity

It is expected that learning resources will tend to have a hierarchical structure where, for example, multiple learning objects are joined together in a module, which is part of a larger aggregation, such as a course. Furthermore, as discussed so far, the introduction of comprehensive metadata will improve the findability and reusability of learning resources, thus leading to scenarios where a given learning resource is referenced by multiple, otherwise independent, resources in the hierarchy. Such data cross-referencing is expected to be done using URLs which point to information hosted at various locations across the internet (for example using the proposed "isPartOf" and "isBasedOn" metadata fields). It is natural to expect that these URLs will decay over time and some of them might become

unavailable, either due to the resource being moved, expired, deleted, or corrupted.

The concept of metadata longevity [R39] is based around the idea that the existence of the metadata needs to be ensured even in the absence of the original data to which it was originally assigned. By decoupling the metadata from the resource itself, it is possible to provide descriptors of what the original data was, and to assist in its interpretation, even when the original is not present. FAIR providers are encouraged to define a metadata longevity plan, fulfilling those objectives.

4 FAIR-by-design learning materials creation

To ensure consistency in the process of preparation of FAIR learning materials, a methodology that will act as an overarching strategy for implementing a FAIR-by-design approach is necessary. In this way, a set of guidelines can be defined to systematically approach the development of FAIR learning materials.

The proposed methodology outlined in this section builds on the previous work related to FAIR training materials, in particular [R12], as well as the OER initiatives, aiming to take advantage of current best practices, experiences, guidelines and other acquired knowledge. It targets all activities related to the development of new learning materials by formally augmenting the traditional lifecycle of learning materials development (i.e. instructional design process) with additional aspects that will ensure the FAIRness of the resulting output.

While this methodology is to be used by instructors, in particular instructional designers who prepare the learning materials, its outcomes (FAIR learning materials) are to be FAIR from both the instructor and learner perspective. That means that the FAIR learning materials are to be:

- Easily findable by prospective learners (in the final consuming format) and other instructors (in the raw editable format)
- Accessible by learners and other instructors with all necessary descriptions and details available at the point of access
- Interoperable in the sense of usable for consumption on different platforms for learners, and provided using standardised metadata, formats and tools to be used by other instructors
- Re-usable by other learners outside the initial target group and by other instructors that would like to design new learning materials based on existing ones

To achieve this goal a number of guidelines and recommendations need to be followed during the design process. A high-level overview of a learning materials design workflow incorporating these FAIR augmented steps is

presented in Fig. 4. The main idea of the workflow is to implement the methodology by paying attention to the FAIR specific actions in each stage of the implementation of the well-accepted backward instructional design process.

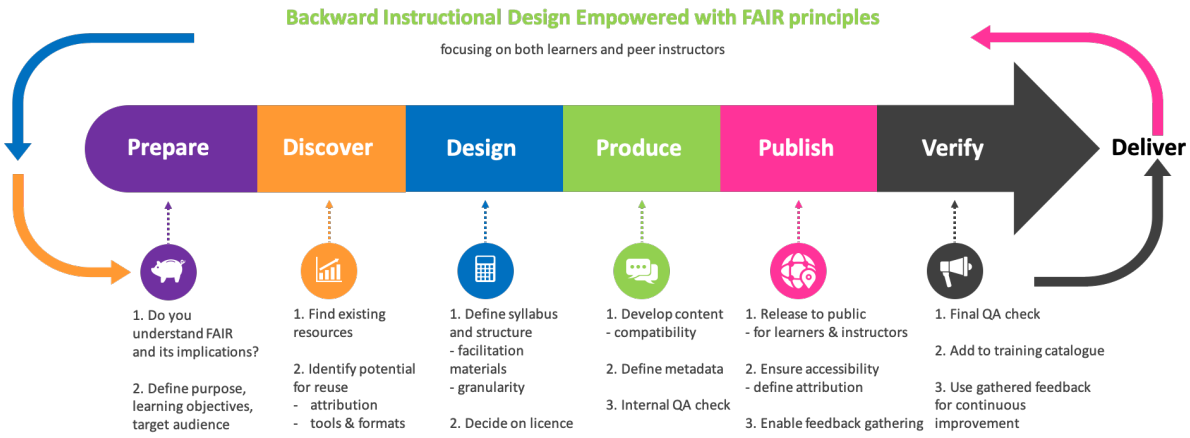

Fig. 4 - FAIR-by-design learning materials workflow

4.1 Workflow stages description

The instructional design and development process starts with a creative idea for new learning content. The typical learning content development process then continues with the analysis, design, development, implementation and evaluation phases focusing on the overall aspects, structure planning, content creation, content delivery and feedback analysis respectively [R40]. For the purposes of the FAIR-by-design methodology for the creation of learning materials, this process is adapted and extended to include additional specific steps and sub-steps that focus on producing FAIR learning materials.

4.1.1 Prepare

Fig. 5 - Goals of the prepare stage

Implementing the FAIR guiding principles for learning objects requires taking into account a number of aspects before starting the actual work on designing and then creating the learning objects (content and accompanying tools and resources). Thus, the first step is to understand what other expertise is required to ensure FAIRness of the produced materials in addition to the traditional learning materials production proficiency. Or, in other words, what are the specific skills that need to be obtained in order to successfully proceed with the instructional design of FAIR learning materials?

Aligned with the joint COL-UNESCO OER Basic guide to OER, that defines the skills requirements for work in OER [R41], the skills required to start producing FAIR learning materials should include:

- Expertise in advocacy and promotion of the FAIR guiding principles;
- Expertise in curating and sharing FAIR data enabling:
    - Efficient application of the concept of metadata;
        - Choosing and implementing metadata schema with accompanying controlled vocabularies and ontologies;
    - Understanding the concept of storing and indexing learning objects:

        - Use of persistent identifiers (PIDs) for unique identification of the learning objects;
        - Use of learning objects repositories or learning management systems and platforms as the designated stores for learning objects;
- Legal expertise to be able to recognise, define and combine Intellectual Property Rights, licensing, attribution and citing of learning materials and other resources;
- Technical expertise in:
    - Different tools and formats used for creation and delivery of learning content and their interoperability;
    - Use of versioning during the creation and maintenance of different types of materials and resources;
- Good communication, collaboration and research skills that will support the process of co-creation, sharing, finding and reuse of learning materials and implementation of best practices and other related policies;
- Instructional design and development expertise [R42] necessary for the creation of high-quality learning materials.

While the last three points are to some extent "traditionally" required skills [R43] when it comes to instructional design and development, the specific tasks and steps related to these activities need to be adapted and further extended so that different aspects of the FAIR principles are incorporated within.

4.1.1.1 Expanding instructional design with FAIR related skills

Understanding the FAIR guiding principles [R44] is an essential step towards the goal of managing FAIR learning materials. One of the essential concepts of curating and sharing FAIR data is the use of rich metadata description of all resources necessary to support findability and reusability. When it comes to the creation of FAIR learning materials, the designer needs to choose an appropriate metadata schema developed for the purposes of describing learning materials. To ensure maximum compatibility throughout the EOSC community, the use of the RDA minimum metadata schema for learning

resources [R31] is recommended. However, considering that this schema is still under development, one may still opt to use other existing schemas especially if the developed material is specific for a certain discipline, such as the Bioschemas option [R29]. Once the metadata schema is chosen, its related controlled vocabularies should be studied so that the appropriate terminology is used in the design process.

Related to the metadata used for describing learning materials are the concepts of using persistent identifiers (IDs), storing and indexing of learning objects. To ensure FAIRness, it is required that, once the learning objects are created and offered for use and reuse, they should be assigned globally unique PIDs that will provide a long-lasting reference to the digital learning resource. These references should also be used when attributing reused learning materials. One general introduction to persistent identifiers can be found at [R45].

The specific type of persistent ID to be used is very much related to the choice of where the learning materials are going to be stored and offered for access to learners and instructors. There are multiple choices available including general data repositories, learning objects repositories and/or learning management systems and platforms. These can be institutional, project-based, or public, and they can also be focused on generic or specific domain content. Also, it is preferable that the chosen location for storing the learning materials is harvested (manually or automatically) by a relevant training catalogue. Having a catalogue entry greatly increases the findability and reusability of the learning material while reducing the concept of importing the same learning materials in multiple repositories or learning systems. Finally, it is good practice to make the choice of the location where the generated material will be stored in advance, as the destination may impose limitations to the type of materials supported, formats and tools used, etc.

The development of legal expertise in concepts such as Intellectual Property Rights (IPR), licensing, attribution and citing is another essential skill that needs to be acquired. The importance of understanding and applying these concepts is twofold: they play a major role in the process of selecting existing learning materials that can be reused during the creation of new learning

materials, but also they are needed so that the newly created learning material can be offered for reuse to others. Some of the most important aspects that need to be clear at the beginning include:

- What is IPR [R46]
    - Intellectual property refers to the creation of intellectual activity and IPR protects the interests of the creators and owners by providing them with rights over their creation.
    - When it comes to the creation of learning materials, the copyright and related rights branches of IPR are used, defined to protect, i.e., literary and artistic creations, performances, phonograms, and define the authors', owners', performers', producers' and broadcasters' rights [R47].
    - Note that in the case of management of learning resources repositories, other related exclusive rights may also be of interest as is the case of Sui Generis Database Right (SGDR), which is different from the copyright protection granted to databases. According to the Directive 96/9/EC on the legal protection of databases, copyright protection will be granted to those databases which "by reason of the selection or arrangement of their contents, constitute the author's own intellectual creation" (art 3(1)). On the other hand, regardless of copyright protection, a database may be protected under the SGDR if it "shows that there has been qualitatively and/or quantitatively a substantial investment in either the obtaining, verification or presentation of the contents" (art 7(1)).
- How copyright defines exclusive rights and free uses
    - Original work can be protected by copyright law that grants the owner exclusive right to control certain rights such as reproduction. The copyright is owned jointly by all authors, or it may be owned by the employing institution.
    - Generally, use of copyright protected work requires permission from the owner. In absence of exceptions or limitations, one can reuse an existing work if it is licensed to the user or it is licensed to the public

using a public licence such as the Creative Commons (CC) licences [R48] or Free-Libre / Open Source Software (FLOSS) [R49] licences.

- How to use licensing as a tool to enable reuse
    - One of the best sources of information for licensing reusable learning materials is Creative Commons. Their Licence Chooser tool [R50] helps authors share their work in a standardised way providing copyright licences that enable sharing and reuse of the creative work under the chosen conditions. The available CC Licence options can be found at [R51]. It is recommended that the least restrictive CC BY licence, requiring only that credit is given to the creator when reusing, is used when creating new learning materials.
    - It is essential to understand that when reusing existing learning materials, one must ensure that the licences of the included and adapted materials are compatible with each other. For these purposes, the CC licence compatibility chart can be used [R52].
- How to use attribution and citing
    - The right to attribution is a moral right of the authors that protects the personal relationship between the author and the created work even if the creator does not own the copyright.
    - Acknowledgement of the reused materials through attribution is always strongly recommended (even if it is not a requirement of the licence). On the other hand, one condition that is required for all CC licences is attribution. The ideal attribution should include the title, creator, source and licence. For more detailed instructions Creative Commons offer a wiki page with the Recommended practices for attribution [R53].
    - Citing can be used for including and referencing restricted works with limited copyright. However, in the case of using direct quotations with citing, it is essential that the amount of information referenced is very limited. In addition, it is recommended to quote works that were already made available to the public in a lawful way and, when possible, to provide the original source and the author's name.

        - Depending on the country, the reproduction and communication of a protected work may be carried out for the sole purpose of illustration for teaching or scientific research, as long as the source, including the author's name, is indicated, unless this turns out to be impossible and to the extent justified by the non-commercial purpose to be achieved [R54].
    - Note that another moral right is the right of integrity which provides the author with the right to object to any modifications of the work that can be considered as prejudicial to the authors' honour or reputation. Thus, even when reuse and modifications are permitted, they should be done in such a way that does not include a derogatory treatment of the work.

4.1.1.2 Ideation of the FAIR learning materials

In addition to these newly developed skills, best practices for the instructional design process [R55] emphasise that, in the initial preparation phase, there should be a clear definition of the overarching aspects and considerations related to the learning materials that are going to be created. In the FAIR-by-design methodology, these aspects should be expanded with additional considerations that might affect the FAIRness of the produced result:

- What is the purpose of the learning materials? This includes when and how the learning materials can be used and for what purposes they were initially developed, also known as primary use.
    - Once defined, the primary use may impose restrictions to the type, tools and formats of the created learning materials including the existing learning materials that might be reused (i.e. e-learning interactive content might be unsuitable for face-to-face or webinar learning settings). In some cases, the primary use might identify the delivery platform for the learning material in this initial preparatory step. If possible, the choice of a delivery platform should be postponed to the design step as to ensure that the choice is

compatible with the types of learning objects identified for potential reuse.

- What are the learning objectives? What competences will be gained after successful completing of the learning process?
    - To ensure standardization and wide understanding of the defined learning objective, it is best practice to define the learning objectives using a well-known taxonomy such as the Bloom's taxonomy [R56]. This approach will significantly improve the findability of the produced resources, as well as the potential reuse of individual learning objects in different aggregations.
    - While traditional competences indicate what a person should know at the end of a study programme, the concept of microcredentials is used to certify the learning outcomes for short-term learning experiences. Thus, if applicable, it is recommended that the microcredentials are also defined at this stage. For more information about microcredentials please refer to the work of Task 2.4 in the Skills4EOSC project. In this way, it can be clear how the specific set of learning materials fit into a larger skillset, such as the minimum viable skillsets and profiles developed by Task 2.1 in the Skills4EOSC project.
    - Note that to ensure the highest degree of reusability, each reusable learning object should have one well-defined learning objective. Multiple learning objectives should ideally be broken down into multiple learning objects.
- Who is the target audience? - is there a primary audience and is there anything specific that needs to be taken into account, such as localisation to cultural context or native language
    - The description of the target audience should follow the rules of the controlled vocabulary of the metadata schema that is going to be used. This approach will ensure consistent description of the learning material improving findability for both instructors and learners.

- Note that at the moment the RDA minimum metadata schema does not offer a controlled vocabulary for these purposes but does recommend its use [R37].

- What is the overall scope of the learning materials? - is it going to be a single learning object, or an aggregation of some sort such as a course, or maybe a learning path
  - The overall scope of the learning materials defines its granularity and has a direct effect on its reusability. Ideally the creation of learning materials should be on the level of a learning object so that they can be reused as flexibly as possible.
- Are there any prerequisites? What does the audience need to know or understand before starting the learning process?
  - The prerequisites help position the learning materials in a wider context and define their place in a higher-level aggregation such as a learning path. Using a standardised, well-adopted vocabulary to describe the prerequisites helps other instructors better understand the learning content and how it can be fit together with other learning objects.

4.1.2 Discover

Fig. 6 - Main goals of the discover stage

Once the preparation phase has been completed, but before the design and development of new learning objects is undertaken, it is a recommended best practice to discover existing related learning materials. Depending on the subject area, vast amounts of existing learning resources might already exist. Such reusability is at the core of the FAIR principles, promoting the extension and improvement of existing work.

During the discovery process, it is expected that not all found material will be open and freely accessible. Should such non-open resources need to be incorporated into the design of the new learning objects, care must be taken to ensure that the material is at least FAIR, before making the final decision. In this context, it is very important that clear accessibility information should be provided by the resource, outlining its access rules and criteria.

Recognizing the real-life benefits of sharing learning resources, a number of initiatives currently exist which facilitate the exchange of Open Educational Resources (OER). These initiatives range from public digital libraries dedicated to OER content, to standalone academic institutions with (in)formal policies on OER content distribution, and specialized search engines. OER Commons [R57] is one such digital library which is built using the community model, where anyone is free to both download existing resources, as well as submit new ones. The USG (University System of Georgia) [R58] and the OpenMichigan (University of Michigan) [R59] portals are examples of institutional repositories for OER produced as part of the curricula in the respective universities. The majority of the content found in these institutional repositories is reuse friendly, licensed under a permissive Creative Commons licence. Finally, dedicated OER search engines have also been developed, which scour multiple digital libraries and institutional repositories for OER which match a given, user supplied, criteria. Examples in this area include the Mason OER Metafinder [R60], the MERLOT search engine [R61], OASIS [R62], and OERTX [R63]. Various different types of learning resources are indexed, such as: videos, podcasts, complete courses, digital textbooks, course modules, open access books, and supporting course materials. Even though not all indexed content might be open, search results

can be limited through the use of explicit filters targeting specific licenses and access policies.

One of the added benefits of using digital OER repositories is the fact that they usually associate valuable metadata to the published learning resources. This metadata, apart from being related to the content at hand, can also serve an additional purpose, that of discovering related material. As mentioned previously, multiple metadata schemas, including the RDA schema extended with the proposed changes, include fields that can be used for content discovery, such as the "isBasedOn" and "isPartOf" fields. One such real-world example is the training portal of the Galaxy Project [R64], which uses the BioSchemas' TrainingMaterial profile [R65] to describe available resources. This profile contains the "isPartOf" field, allowing learning resources to specify the learning unit that they are part of [R66].

Valuable learning materials can also be found in more general-purpose repositories, not intrinsically related to OER. Zenodo [R67], [R68] is a generic example which is also commonly used for publishing learning resources, with one reason being the allocation of a unique persistent identifier to each uploaded item in the form of a Digital Object Identifier (DOI). Additionally, dedicated search tools that can query thematic repositories which do not necessarily host learning resources themselves but might contain content beneficial to the development of learning resources can be exploited as well. FAIRsharing [R69] is an example of such a search engine, which indexes FAIR-friendly databases.

Another valuable source of learning material which can be reused is the output of European research projects and initiatives. Many such projects, across different scientific disciplines, have created and published, under permissive licenses, material which can be incorporated or used as basis for future content. This material can either be found using dedicated search engines targeted at a specific scientific domain, or in certain cases, dedicated e-learning platforms that have been created as part of the project itself. The NI4OS Training Platform [R35], OpenPlato [R33], and EOSC Pillar [R70] are such examples, primarily focused on the topics of open science, research data management, and FAIR practices. GoTriple [R71] is another example of

a search engine, but in this case dedicated to a different scientific area, that of social sciences and humanities. It can be used to find relevant research data, publications, and projects. For life sciences, the ELIXIR TeSS Platform [R72] provides various relevant learning resources, including courses, videos, presentations, and handbooks. An alternative for physical sciences is the PaN EU Training platform [R73]. For material primarily related to computer science topics, the EOSC-Synergy Training Platform can be used [R74]. EOSC-Synergy also includes introductory courses to popular computer science concepts, tools, and services which can be of use in other subject areas as well.

No matter the source of the content, careful attention should be given to the associated metadata, licensing information, and their impact on the reuse and modification. The mentioned search engines and repositories either mandate content to be openly accessible and licensed under a permissive Creative Commons licence or contain explicit licensing information and access restrictions.

4.1.3 Design

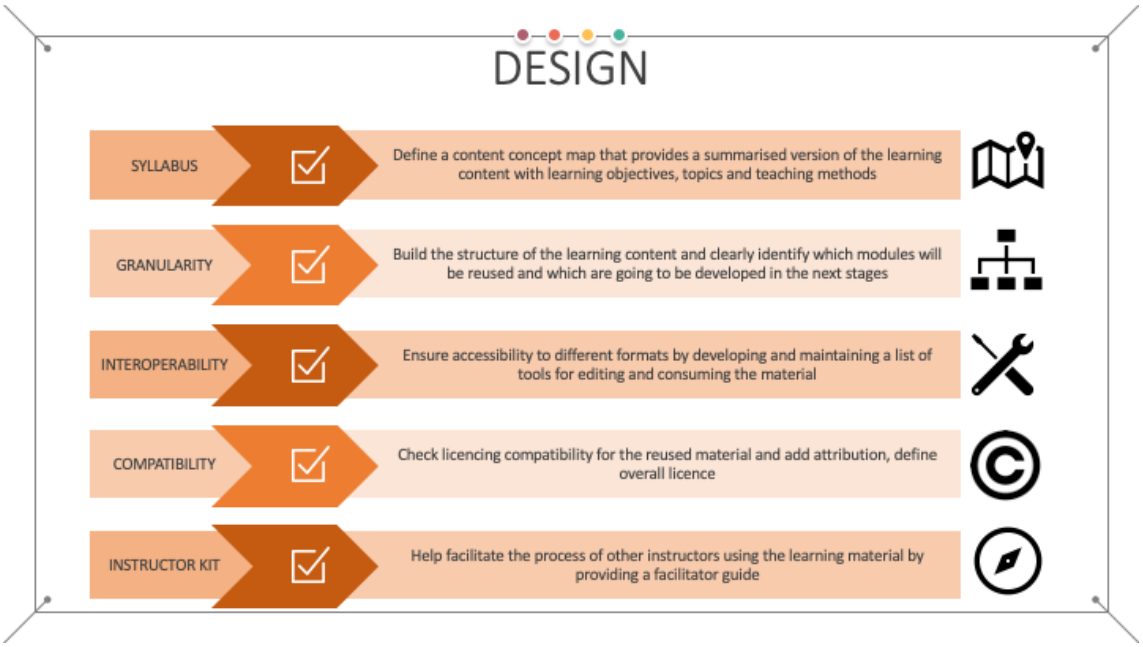

Fig. 7 - Main goals of the design stage

Once the learning objectives and target audience are defined, the discover step provides an opportunity to explore how learning materials with the identified or similar learning objectives are designed and implemented. This exploration enables the designer of the instructional materials to take into account different approaches and methodologies which:

- can serve as an inspiration,
- can be used as additional or supporting materials, or
- can be reused, in part or as a whole, as the main resources that will help learners achieve one or more learning objectives.

The choices made need to be aligned with the plans on implementing assessment and evaluation, as they influence the selection of teaching methodologies and content type, making the design outcome oriented. An optimal strategy is to aim for building in a variety of different assessment techniques such as discussions and reflections, quizzes, simulations, projects, case studies, self and peer evaluations, etc.

Following the stages of the backward instructional design process, during this third step of the methodology the syllabus, needs to be developed [R75]. The main goal of the syllabus, also known as the content concept map, is to provide a summarised version of the learning content, listing fundamental information with clear learning objectives, modules (or topics) that will be covered, and teaching methods employed. The programme needs to be tailored according to the needs of the identified target audience and aim. Best practices [R76] are to facilitate a rich learning experience by defining a modular structure that will be composed of the essential knowledge necessary to achieve the learning objectives augmented with additional resources for further elaboration for learners that are interested in obtaining more in-depth knowledge on a given topic. In summary, the course syllabus is a document that presents the purpose of the learning content with precise description of what is to be learnt, how and when under the assumption of the defined prerequisites.

The defined syllabus can then be used as a blueprint to build the structure of the learning content and clearly identify which modules will be reused

based on the output from the previous discovery step and which are going to be developed in the next stages. When building more complex, aggregated learning content, they should be organised in a corresponding group of smaller aggregates that decompose to the level of learning objects (Fig. 8).

The overall structure of an aggregated learning content, i.e. course or training, should include:

- Sections - each referring to one overarching theme, which are comprised of
- Modules - each with defined main goals, description and rationale, which are comprised of
- Learning units - each corresponding to one lesson with its specific objectives, activities and tasks, description of organisation and further reading, which are comprised of
- One or more learning objects - each with a well-defined learning objective.

Each module in this organisation should end with a module summary and a wrap-up activity, i.e. reflection and assessment, that will enable the learner feedback on whether the module outcomes have been achieved.

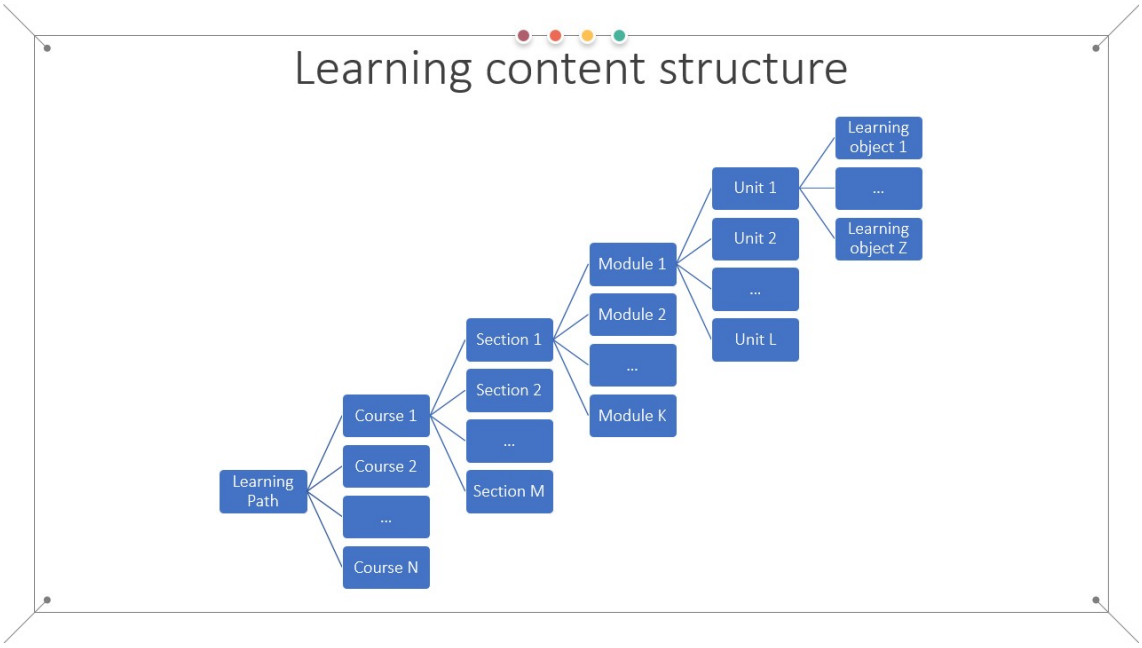

Fig. 8 - Hierarchical structure of aggregated learning content

From a FAIR perspective, the adoption of existing learning materials in the final structure at this step will depend on:

- Granularity of the materials - if the material represents a higher-level aggregation it will be more difficult to incorporate into a specific structure. In this case, it should be investigated if it can be used in part with some modifications. This entails the requirements that the content is provided under a licence that allows derivations and modifications and having the appropriate tools and experience for editing the content.
- Interoperability - the selected existing learning materials may be provided in different formats and various tools may be needed to consume their content. The final choice must be done so as to ensure that the combined materials are interoperable and can be consumed by the learners. To achieve this, a list of all tools necessary to access the learning materials needs to be developed, ensuring that each tool on this list is accessible to the learners via the chosen method of delivery. Similarly, a list of tools for editing the materials need to be maintained for future reusability purposes.
- Adaptability - it is preferable to choose existing learning materials that are adaptable to various learning and technical contexts as they provide flexibility when reused in an alternative context or environment. This entails that the material will still make sense and fit well in other structures, not just the original structure and context for which it was produced.
- Licence compatibility - All chosen materials for reuse should be available under a licence that permits reuse in the way it is planned to be used (see Fig. 9 modify and adapt): for commercial use or not, with modification and adaptation permission or not.

When the final choices on reuse are made, the attribution for each of the reused learning materials should be defined and added to the information in the appropriate structure level (section, module, unit). This will not only ensure that licences are respected where attribution is required, but it will also promote transparency and ethical conduct by providing attribution.

It is important to note that at this stage, the list of selected materials for reuse will also influence the overall licence of the produced FAIR learning materials. Thus, it is recommended that all licences are carefully checked,

and an overall licence for the FAIR learning materials is chosen so that is aligned with the original licences of the reused materials as some licences do not allow changes (Fig. 9). When it comes to the Creative Commons Licences, this means that special attention is needed when the SA (Share Alike) configuration is present, restricting the use to the same licence. Another observation is that the ND (no derivative works) cannot be combined with SA, as SA applies to derivative works. Creative Commons provide a very good tutorial on the topic of remixing CC licensed work [R77].

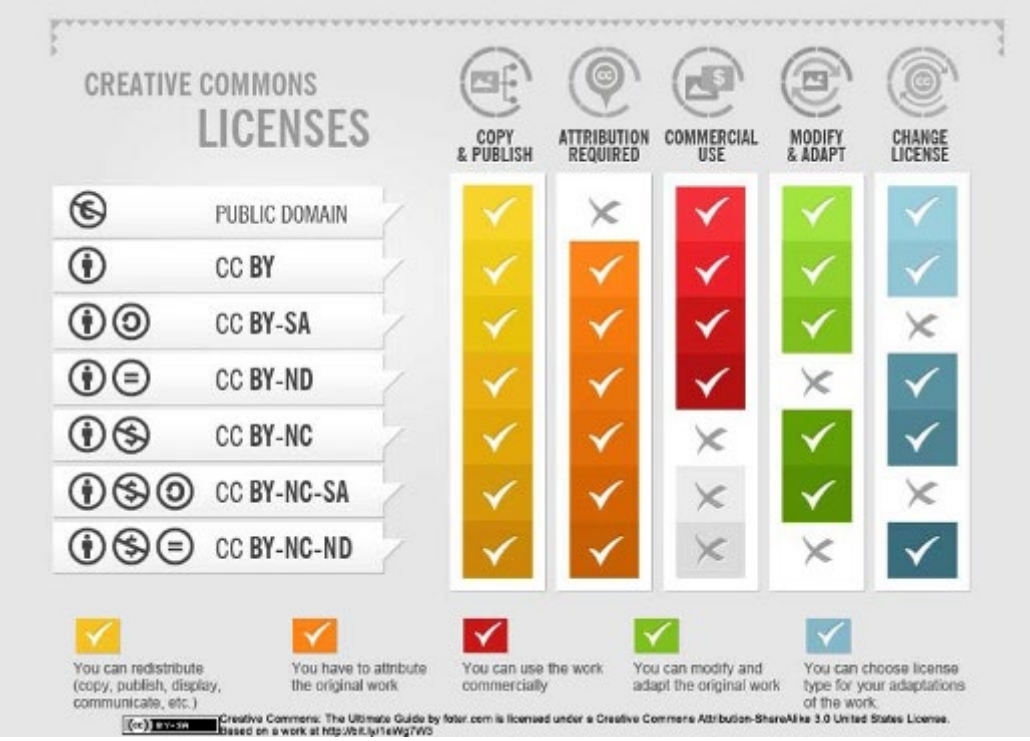

Fig. 9 - Comparison of different Creative Commons licences

Copyleft licences [R78] are another set of licences that provide the freedom to copy and share the work with others, together with the freedom to modify the work and distribute modified. However, in this case the work can only be distributed under the same or equivalent licence. Copyleft is mostly used for software, but it can be used on any type of work. Table 1 provides a generalised summary of the licensing specifics and restrictions of different types of licences that can be used for learning materials.

Table 1 - Generalised summary of different licenses (based on [R79])

	© Copyright	Copyleft	cc Creative Commons
URL	https://www.copyright.eu/	https://copyleft.org/	https://creativecommons.org/
What is a user allowed to do with the work?	What author/owner dictates	What user wants under certain rules	What user wants within the licence restrictions
Clause of the use	As author/owner dictates	requires attribution to author and copyleft	Requires attribution to author
Re-licencing	As author/owner dictates	Derivative work cannot be released as proprietary and should be licensed under a copyleft licence	Derivative work can be released under another licence or as proprietary (as long as the share alike rules don't apply)
Commercial use	As author/owner dictates	Permitted when using the work as a commodity tool or component to provide a service or product	Permitted for certain CC licence types (as long as the non-commercial rules don't apply)

Note: it is strongly recommended to always check the terms of the specific licences, because they may provide additional/other obligations regarding the items addressed in this table.

To be effectively reused, the learning materials should also be augmented with accompanying materials that comprise the so-called instructor kit, or facilitator guide [R80]. The kit is especially important for traditional learning settings (face to face), but also for hybrid and blended delivery of the learning material. Even in the case of online learning, there usually is an option for some type of interaction with the instructors, and the way this interaction is organised, its timing and frequency, scope and methods should be explained in the accompanying material.

The instructor kit should help facilitate the process of other instructors using the learning material. Thus, a comprehensive instructor kit should contain the following information:

- How the material is structured in different sections, and what should be considered before starting and after completing each section (for example, ice breakers and reflections);

- Tips and tricks on how to make the delivery more effective based on the proposed teaching methodology;
- Best practices on the agenda (timing, pace, breaks) while covering the content;
- When to start and how to manage discussions;
- Organising and running different exercises;
- Room preparation and set-up, tools and props required (in case of physical delivery);
- A workbook that can be shared with the learners to be used as a study guide;
- Script/Notes that provide instructions what needs to be covered in each section;
- Assessment setup and questions;
- Step-by-step instructions on running exercises for a particular content together with handouts that should be provided to the learners;
- Other resources, such as feedback form or template certificate, attendance forms, etc.

An example comprehensive facilitator guide on how to organise face-to-face training events covering various logistics aspects is the TRIPLE Training Toolkit [R81].

4.1.4 Produce

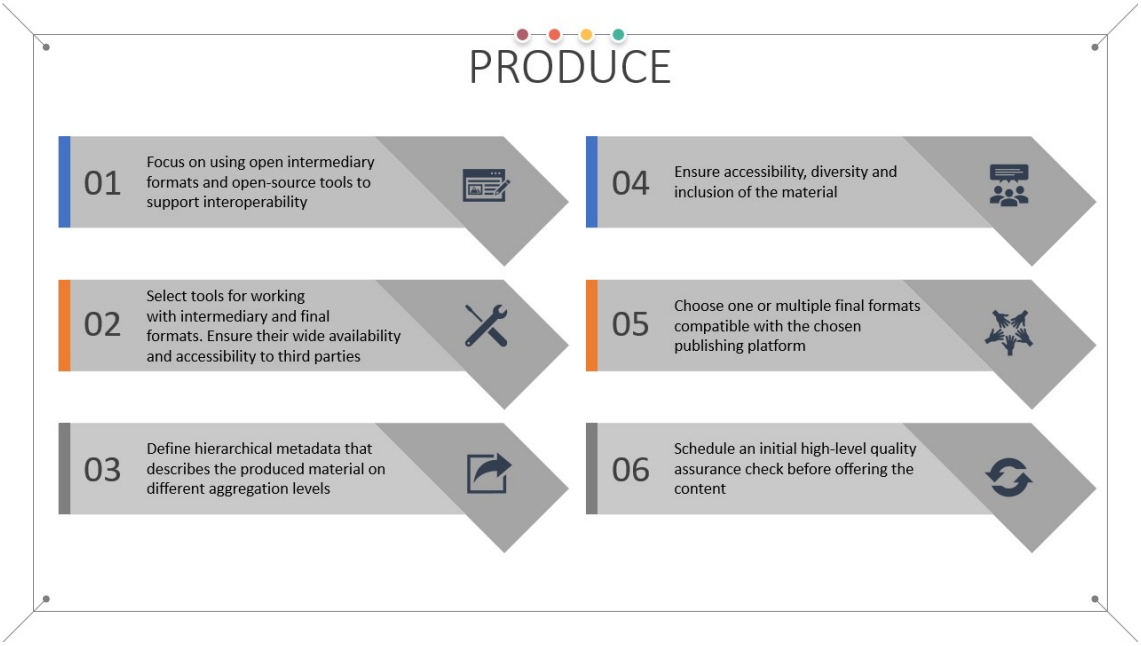

Fig. 10 - Main goals of the produce stage

The next step after creating the syllabus and defining the overall structure of the course is to develop the new learning objects themselves. Throughout this activity, there are three important aspects which need to be taken into consideration:

- The intermediary and final file formats of the produced learning objects;
- The required software tools for producing and consuming the learning objects;
- Ensuring future-proof compatibility and longevity of the learning objects.

In the subsections that follow, it is discussed how each of these aspects relate to the FAIR principles. A taxonomy of file formats for learning objects, along with examples of tools which can be used to either create or consume the output is also provided to further support the development activities.

4.1.4.1 File formats for learning objects

Selecting the appropriate file formats for the learning objects is very important and can have a large impact on the overall experience that the end-users have during the learning process. As new, feature-rich, and

interactive forms of content appear, it is a common occurrence for the final output format to be different than the intermediary one. The intermediary file format is the format which is used during the development phase of the learning objects. Once completed, additional tools can be used to convert this intermediary format to its final representation, which is most often no longer manually editable. Both the intermediary and final formats need to be chosen with care, since a restrictive or proprietary choice might limit the number of people who can contribute to the content during its development or make the final output inaccessible due to the lack of supported software client applications.

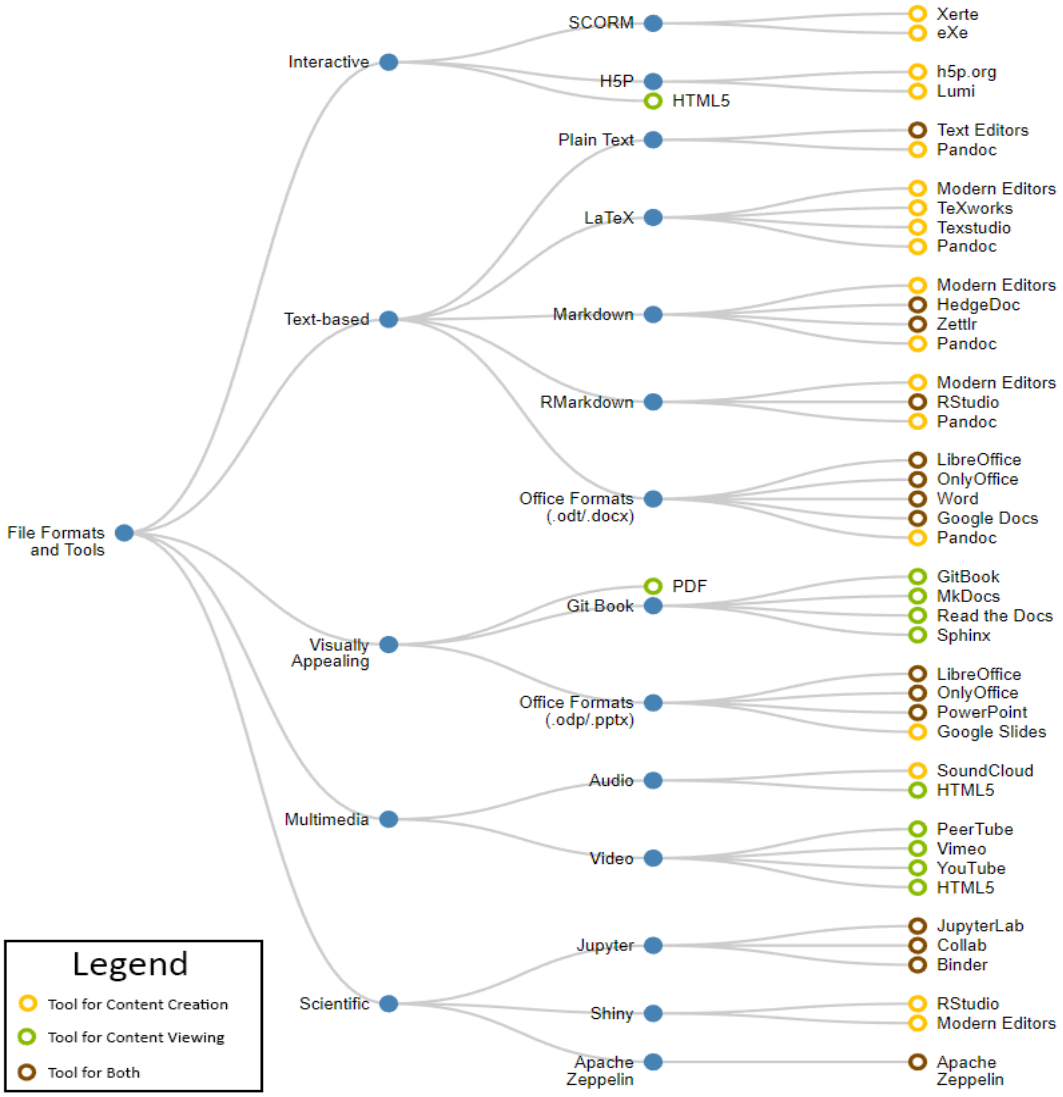

Fig. 11 - Taxonomy of file formats and tools

Depending on the nature of the learning objects, their domain area, and the technical proficiency of the content creators, a number of popular file formats exist today, such as: SCORM, H5P, various text-based formats, PDFs, multimedia, scientific notebooks, or even complete e-books accessible on-the-go. To better visualize the available options, Fig. 11 presents a taxonomy of file formats along with potential tools which can be used either for development or consumption. The file formats are divided into 5 distinct categories, based on their characteristics, interactivity level, content type and domain:

- Interactive,
- Text-based,
- Visually Appealing,
- Multimedia, and
- Scientific.

The interactive category is comprised of three file formats: SCORM, H5P, and HTML5. Both SCORM and H5P are dedicated file formats for learning objects and enjoy wide ranging support from learning management systems (LMS) today, making them a popular choice. The main benefits of using a file format from this category is the high level of interactivity that can be achieved, since various activities can be directly embedded in the content itself, such as quizzes, interactive maps, videos, audio, and animations, thus providing a cohesive, all-in-one experience to self-paced learners in an asynchronous learning environment. With advancements made in terms of web standards in recent years, all modern web browsers support the consumption of both SCORM and H5P resources, since their final representation is HTML5, in the majority of cases. Technically proficient educators can also opt to develop their learning objects natively with HTML and other relevant technologies, such as JavaScript, in cases where the SCORM or H5P suite of tools do not meet their demands. Such manual development of learning objects without using an existing framework is rarely seen in practice and is generally discouraged due to the increased time requirements and complexity. However, it must be noted that as meaningful changes to the final output of both SCORM and H5P cannot be made easily, instructors should make the

intermediary representation of these resources publicly available as well, allowing others to reuse parts of the content more easily.

The second category in the taxonomy is comprised of various formats with a common characteristic - they are all text-based and can be easily edited with regular text editors, in addition to more feature complete software dedicated to a particular file format. Plain text files, while being easy to write and consume, are less visually appealing since formatting and typography customization are limited. On the other hand, virtually all computing devices ever produced have support for consuming basic plain text content. The next three formats: LaTeX, Markdown and RMarkdown are intermediary formats, rarely used as the final output. They combine the versatility of plain text with advanced formatting features and can be converted to various representations (e.g., PDF, HTML) with converters. These new formatting features are introduced into what are otherwise plain text files through the use of a specialized syntax with variable complexity, with Markdown being least complex and LaTeX being the most complex. All three formats are in wide use today among educational communities, with the main benefits being the easy conversion to a web-based representation, the ability to write in plain text without specialized software, and easy file versioning. Even though reverse conversion from HTML and PDF back to the intermediary format is technically possible, it is still advised to publish the source text material as well. Finally, the office formats such as the open source .odt and Microsoft specific .docx enjoy the largest popularity today, as a result of modern, well-tested text processors which provide graphical user interfaces, eliminating the need for specific text-based syntax in order to achieve the desired representation and formatting. While it is possible to use both HTML and PDF as the final representation in this case as well, Office formats are more limited in terms of source code representation (important for specific disciplines) or embedding of third-party content (increased interaction).

While it is true that visually appealing material can be achieved using a wide variety of file formats, this category introduces 1 intermediary format and 2 final formats which are known for the pleasant visual properties of their content. Slides or presentation file formats such as the open-source .odp or

Microsoft specific .pptx not only offer creation of text-based presentations, but also allow the introduction of additional material, such as interactive objects, video, audio, and animations. With the use of software addons, it is also possible to convert such media heavy presentations directly into HTML5 or SCORM, instead of static PDFs.

Even though static, PDFs are widely used as a final format in education communities because of their uniform representation, independent of the device. As discussed previously, all text-based intermediary file formats can be easily represented as PDFs, preserving their formatting. Directly editing PDF files or reverting the PDF conversion process, while possible, is usually discouraged, due to the requirements of specialized software and potentially inconsistent output results. PDF files can also be digitally signed and optionally protected in a standardized manner, guaranteeing their authenticity and integrity.

For large amounts of mixed content that includes both text and multimedia, and which is constrained by the sizes of physical paper formats enforced by PDFs, Git books are a feasible alternative. Git books are most often represented as complete web sites, comprised of multiple HTML pages, automatically rendered via a conversion process which takes one of the plain-text formats as input. Ready-made frameworks exist for the development of such Git books, taking care of their overall design, user interface, and even publishing. The term Git in the name of the format refers to the source code management system of the same name, which is most often used for tracking changes made to the content of the text-based intermediary formats among several contributors. It is a common practice to make the Git repository hosting the intermediary text files public, promoting collaboration, and allowing external contributions to the work. The Open Science Training Handbook is one such real-world example of a Git Book which is publicly accessible as a set of nicely formatted web pages [R82], while keeping the source Markdown files open as part of a Git repository for anyone to edit and contribute additional content [R83].

Multimedia resources can either be embedded directly into the learning objects, in case an interactive final file format is used, or can be posted as

independent files, linked to the main content using relevant metadata fields and references. The main challenge faced when dealing with these file formats is the expertise required for their editing, the need for specialized software tools, and the hosting location. Multimedia files require much more storage space compared to the other file formats, which usually leads to them being offloaded to external, third-party hosting platforms. In such cases, when resources are detached from the main body of content and hosted on third-party platforms, special care needs to be paid to the terms of use of the third-party services, and their reliability. It is recommended for a link-rot strategy to be in place, dealing with the problem of hyperlinks which are no longer accessible after a period of time due to them being moved to a new location, deleted, or permanently made inaccessible for any other reason. It is advised to monitor for such occurrences of link-rot, and overcome them by relinking or reuploading (if the licence allows) any missing content which is no longer accessible.

In certain cases, it might be beneficial to allow learners directly alter the content of the training material, for example during interactive exercises or analysis of results. Interactive notebook formats have rapidly gained popularity in recent years, especially in subject areas that rely on extensive visualizations or scientific data analysis. Jupyter, Shiny, and Apache Zeppelin are all representatives of interactive notebooks, which allow mixing of static and dynamic content together. Usually, the static context is text written in either plain text, Markdown, LaTeX, or RMarkdown, while the dynamic content is represented by statements written in a supported programming language. When consuming the content, users have the option of directly altering the dynamic content, either by modifying the existing visualizations, adding/removing data, or performing further analysis using the syntax of the supported programming language by the notebook instance. It is also possible to make a static export of the notebook content in one of the supported final formats such as PDF or HTML. During the export process the dynamic content is evaluated and its results are statically included in the output, barring future changes, making the material more accessible on

portable devices which might otherwise encounter problems when loading the more resource intensive, albeit interactive, notebook environment.

4.1.4.2 Tools for creating and consuming learning objects

Throughout the learning object development process, attention should be aimed at choosing the most suitable intermediate and final formats for delivery of the given learning material, instead of focusing on a particular tool. Nevertheless, Fig. 11 presents concrete tools which can be used for creating and consuming all of the previously discussed file formats.

The tool selection in the figure is by no means exhaustive and should only serve as a starting point in the decision-making process regarding tooling. Most of the included tools are open-source software with permissive licenses, with notable exceptions being made where the popularity and wide-spread usage of a given software or platform could not be ignored, thus warranting inclusion into the list. This is the case for the majority of multimedia tools and platforms, since as discussed previously the overhead of hosting such files together with the rest of the learning objects incurs a high overhead in terms of compute resources (YouTube, Vimeo, SoundCloud). The list also includes the Microsoft .docx and .pptx formats, as well as the proprietary Google Docs and Google Slides formats, which were included due to their widespread usage and popularity today. Cloud hosted tools supporting these formats are very popular among educators and researchers today due to their effortless collaboration features and general ease-of-use. However, it should be recognized that such convenience comes at the cost of a vendor lock-in to an extent, limiting the interoperability options with other open source tools. Even though there are other alternative office document formats, together with open source office suites, usually in practice compatibility issues arise, and testing should be done on a case-by-case basis.

When it comes to the vast landscape of software tools for editing text-based formats, such as Markdown and LaTeX, applications which offer on-the-fly preview of the written content can provide a more pleasant first user experience and a gentler learning curve. Examples include the Zettlr [R84]

and HedgeDoc [R85] editors. On the other hand, more advanced users might appreciate greater customizability usually attributed to command line utilities such as Pandoc [R86], which can be used as a general-purpose converter to/from various different formats, including producing PDF and web page representation of existing Markdown and LaTeX files.

4.1.4.3 Ensuring compatibility and longevity of the learning objects

Both compatibility and longevity play an important role in the development of FAIR-by-design learning objects. Instructors should strive to use open formats which are tool agnostic and compatible with a wide variety of existing software, thus avoiding proprietary features, which might also lead to an undesirable vendor lock-in effect. The use of open formats not only makes it easier for learners to consume the final output, but also increases its overall reusability, and future-proofs its usefulness. Designing the training materials using standardized and open file formats, decoupled from a specific software tool or even a particular version of a software tool ensures the longevity of the material, and limits the influence of external factors which are not controlled by neither the instructors nor learners consuming it. In the past many file formats have been tightly coupled with specific tooling, such as in the case of Java Applets or Flash, which led to the inaccessibility of many educational materials constructed using these technologies once they were deprecated.

4.1.4.4 Ensuring accessibility

In this stage care must be taken that the content of the newly developed learning materials is accessible for everyone. Herein accessibility transcends the definition in FAIR and refers to the idea that people of all abilities should be able to access the content [R87]. This includes developing content for people with different learning styles as well as ensuring that the content will be accessible to people with disabilities. In addition, it also refers to the idea that the learning materials should include different viewpoints of the subject matter. In other words, the development process should be done in such a way that the newly created learning objects are accessible, diverse and

inclusive overcoming challenges such as physical impairments, learning disabilities, language comprehension, and other limitations.

The Self-Publishing Guide by BCcampus [R87] provides an exhaustive list of challenges that need to be addressed to ensure accessibility, diversity and inclusion while creating new learning materials. There are many other instructional support pages such as [R88], that provide clear, precise instructions on how to overcome these challenges by paying attention to employing more advanced formatting techniques, special colour palettes, alternative text, captioning, etc. Many of these guidelines are easy to implement, yet they go a long way when it comes to accessibility for people with impairments. Such examples include using different level headings for titles, using built-in list tools, using column headers and row headers on tables, and using math tools instead of images of equations. A very helpful set of Accessibility Handbooks available under a CC licence can be found at [R89].

4.1.4.5 Metadata definition across all learning objects

Once the structure of learning objects defined with the syllabus has been completed, the next step is to apply relevant metadata information. Enriching the learning objects with relevant metadata aids their findability and reproducibility, especially in circumstances where a permissive and open licence has also been assigned. Specific metadata fields can also be used to more clearly specify the wider context of a given learning object, allowing it to be independently shared and reused across different media in varying scenarios.

A very important, albeit often overlooked aspect to keep in mind during the development and verification phases of learning objects is that a careful balance needs to be stricken when it comes to the details and surrounding context of the material. On the one hand, it is desired that learning objects should be as standalone as possible, free of interdependencies. This ensures effortless reuse by educators, allowing higher level learning resources (learning units, courses, learning paths) to be created based on composition of both new and existing lower level (in a hierarchical sense) learning

materials and objects. However, on the other hand, from a learners' perspective, such a flat approach with a limited overall context would pose challenges. Learners would need to resort to manually searching for relevant materials on a given topic of interest in case no higher-level learning resources currently exist or are simply not easily findable due to their limited adoption of the FAIR principles.

An elegant way to strike a balance between the reusability of the material and its attachment to a narrower context, useful to learners in their effort to achieve more advanced and focused learning outcomes, is through the comprehensive use of metadata. Metadata information should not be limited only to atomic learning objects. On the contrary, it should be applied across all levels of the hierarchy, stretching from learning objects, to learning units, to courses, and learning paths. The previously discussed and recommended RDA metadata schema is agnostic when it comes to what level in this hierarchy it is being applied to. This is an important feature which allows the same metadata fields to be reused for different kinds of learning materials across the complete hierarchy, without the need to develop new metadata schemas or exchange, analysis, and verification tools.

The three different categories of fields which comprise the RDA metadata schema - Descriptive, Access, and Educational, along with the proposed two extra fields "isPartOf" and "isBasedOn", facilitate effortless reuse of existing materials by educators, while ensuring that the surrounding context is preserved, comprehensible, and easily accessible for learners. An all-encompassing application of the RDA metadata schema across the whole hierarchy of learning material would ensure that the desired FAIR criteria are met by all associated resources. The "isPartOf" and "isBasedOn" properties can also be used to improve the capacity to model hierarchical relationships, even when sharing learning objects independently of their higher-level elements, preserving the overall context in which they were originally defined.

The use of a standardized metadata schema across the complete hierarchy of learning materials also has the added benefit of making the process of assigning metadata information easier, well-defined, and can even lead to

the development of automation tools capable of prepopulating a subset of the fields by themselves, without human interaction. For example, the following metadata fields can be derived from comprising learning resources, when discussing higher level materials:

- Author(s) - a list of all the authors of the encompassed learning material, optionally extended with the creator of the higher-level resource;
- Language - a list of languages in which the included learning material is available;
- Keywords - a list comprised of all the unique keywords assigned to each of the included learning material;
- Licence - depends on the desired licence for the new higher-level resource, as well as its compatibility with the licenses of included learning materials;
- Access Cost - sum of all access costs for the comprising materials, or a previously determined cost, as determined by the creator of the higher-level learning element. If the final learning materials are not open and for free, note that the use for commercial purposes must be allowed by all licenses, the new licence for the higher-level resource licenses of separate learning objects;
- Learning Resource Type - a list containing the information regarding learning resource types of all included materials;
- Expertise Level - in the case of mixed expertise levels for the included material, the most advanced one should be chosen for the new learning element;
- isPartOf - an automatically set reference to the encompassing resource;
- isBasedOn - list of references to the persistent identifiers of all included learning material.

The remaining descriptive, access, and educational fields need to be manually provided or depend upon rules which might be specific to a given subject area.

The use of automated tools for metadata enrichment can very much alleviate this step and ensure it is not prone to human error. Such automated tools can be used for continuous updating of the metadata, ensuring that the information assigned to the higher-level resources is always in sync with the

changes made to the lower-level materials. Alternatively, automated tools can be used in conjunction with manual enrichment, limiting the syncing only to fields which have not been customized manually by a human operator. However, these tools should be supported by the learning management platform that is used to host the learning materials.

4.1.4.6 Internal quality assurance check

At the end of this stage, the whole bundle of learning materials should be ready to be offered to the learners. Before this is done in the publish stage, an internal quality assurance (QA) check is needed to ensure that nothing has been overlooked and that everything combines together as intended.

Best practices [R90] are to have a separate evaluator that will perform the internal QA and provide feedback: whether the learning materials bundle is ready for publishing or there are some issues that need to be addressed in order to ensure high-quality production of the materials and thus high-level user experience by the learners.

The main criteria that should be evaluated during this stage are a mix of standard QA check for training materials and additional FAIRness related checks.

A high-level internal QA checklist that covers all aspects should include the following:

- Overall design;
- Appropriate topic breakdown and structured layout aligned with syllabus;
- Metadata description for all learning objects (aligned with the RDA recommendations, or another domain specific schema);
- Quality of media in the material;
- Matched level of context to target group, easy to consume and understand;
- Appropriate prerequisites defined;
- Content aligned with clear learning objectives;
- Attribution/Citing of external sources;
- Accessible to consumers using simple, intuitive tools;
- Comprehensive facilitator guide;

- Assessment tasks (types and content);
- Use of controlled vocabularies.

Many education-oriented institutions have their own internal QA processes that should be activated at this step. In this case, it is essential that the FAIR aspects are not overlooked, in particular: appropriate copyright and licensing, attribution and citing, metadata and controlled vocabularies use, and reuse possibilities.

4.1.5 Publish

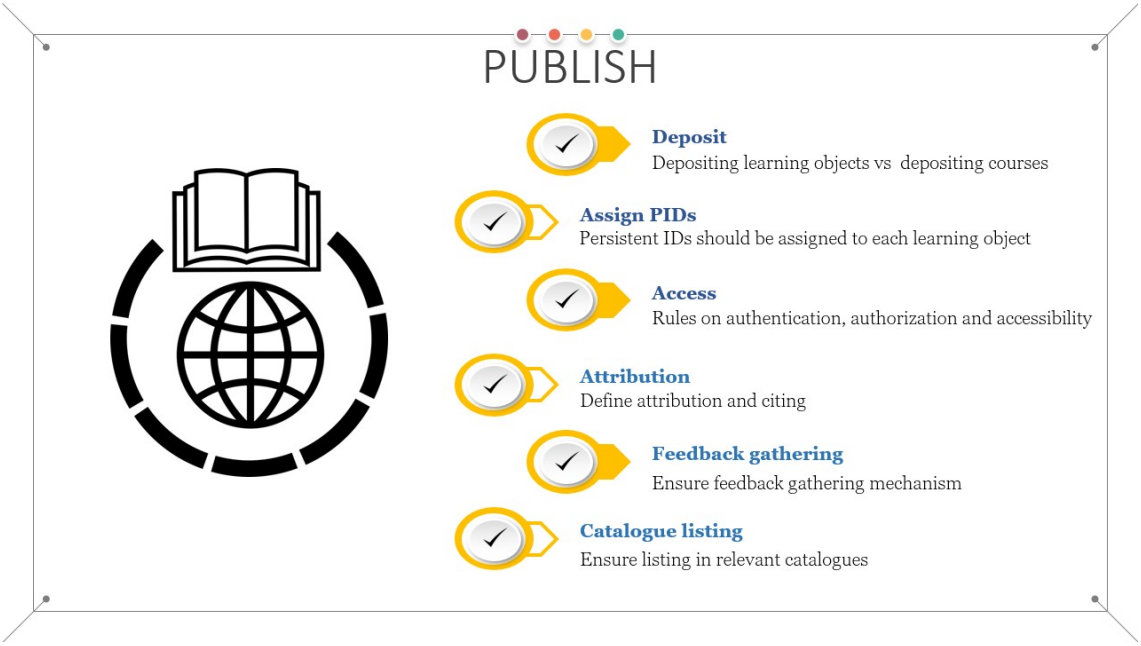

Fig. 12 - Main goals of the publish stage

The publishing phase of the workflow refers to the release of the produced learning objects and associated metadata. The publishing refers to both newly created learning objects and new versions of previously published objects.

A clear distinction should be made between deposited learning objects and learning objects published as part of larger learning units, such as courses. The former should be deposited in relevant repositories [R91], as separate deposits, preferably in source format (editable), while the latter are usually published on learning/training platforms. The main difference between the

two is the target audience. The former target instructors/trainers to use/reuse the objects for producing learning materials, while the latter are mostly targeted toward the learners/trainees for consumption.

Each deposited learning object should be accompanied by relevant metadata, providing information for its discovery, composition, and reuse, and appropriate persistent identifier [R92].

The preferred platform to deposit the editable (source) format of the learning objects should be GitHub. GitHub provides built-in visioning mechanism, enabling the trainers to easily track the versions of the learning objects. One drawback to this is that GitHub does not provide persistent identifiers for the deposited objects. To overcome this shortcoming, the integration between Zenodo and GitHub can be used. Once materials are deposited to GitHub, the authors should navigate to Zenodo, use GitHub credentials for login and using the Zenodo GitHub page, archive the GitHub repository to Zenodo. The detailed procedure is described in [R93]. Once the repository is archived to Zenodo, it will be assigned a persistent identifier (such as DOI).

Given that learning objects should be granular and disassociated from the context as much as possible, as well as to provide easier reuse the deposition in learning object repositories, such as ones given in this list [R94], depositing learning objects in repositories should be done in a flat model, avoiding any hierarchical approach that would potentially hide some learning object within a given context. On the other hand, learning objects that are composed in higher level elements, such as courses, published on learning platforms and offered for consumption by learners, should be done in a more hierarchical manner, depending highly on the context.

Even though the goal of FAIR training material is to be widely accessible, in some cases access rules must be defined, in the form of who has access (authentication) and to what objects (authorisation). Access rules should be assigned consistently. When combining learning objects with different accessibility rules, the most restrictive rules should be clearly stated during the publishing phase. Based on the FAIR principles, it must be assured that the learning materials are "available at the point of access". This means that

they should be searchable and findable, based on the metadata description, and the search result should provide a landing page with all details about the learning materials, including how to get access if they are not open, cost and alike.

At this stage the attribution for the learning materials should be provided so that it can be used if someone else decides to reuse the provided materials. Again, special consideration should be given when simply compiling higher lever learning materials out of existing individual learning objects that have different attributions. In this case, the attribution should note that it refers to the activity of gathering and organising existing materials. In cases when there is a mix of original content and existing learning objects, then the overall attribution refers to the new author(s), while for each reused part clear attribution to the original author should be provided.

Once the materials are published and made available to the public, there should be mechanisms in place that will enable gathering feedback about their use. Collecting feedback on the published learning materials is key to implementing continual improvement. The exact type of feedback collection should heavily depend on the way the learning materials are presented to the users/learners.

In face-to-face (f2f) sessions, feedback can be collected during the delivery of the training and after it. Interactive feedback collection can be more valuable, since the opinions of the learners are collected as they perceive the learning materials.

For online published materials, feedback is usually collected post-festum, after the learner completes a given unit of learning objects.

Another aspect of feedback gathering related to the feedback from the community of instructors/trainers that use the provided learning materials to augment their own trainings or to develop their own learning materials. It is encouraged to enable at least one channel of communication (it can be as simple as a request for feedback via email) that will enable the creator receive feedback from the community on the level of reusability of the offered learning materials. This effort can not only lead to producing higher quality

learning materials by employing the continuous improvement loop, but also to building a strong community of instructors that can work closely together when producing new learning materials.

To make the learning materials more easily findable, after publishing in the relevant repository and learning platform, they should also be listed in relevant training catalogues. The catalogue entry can be done manually, or it can be created by an automated harvester that indexes the particular repository or learning platform. As discussed in the discovery stage, there are different catalogues available: thematic, national, regional, project-based etc. At the moment there is a significant amount of work in the EOSC related projects, led by EOSC Future, aiming to implement training catalogue aggregation of all diverse types of training catalogues into one master EOSC training catalogue [R36]. This type of catalogue aggregation will simplify the findability of training materials for the end-users, giving them (ideally) a single point of access to various learning objects.

4.1.6 Verify

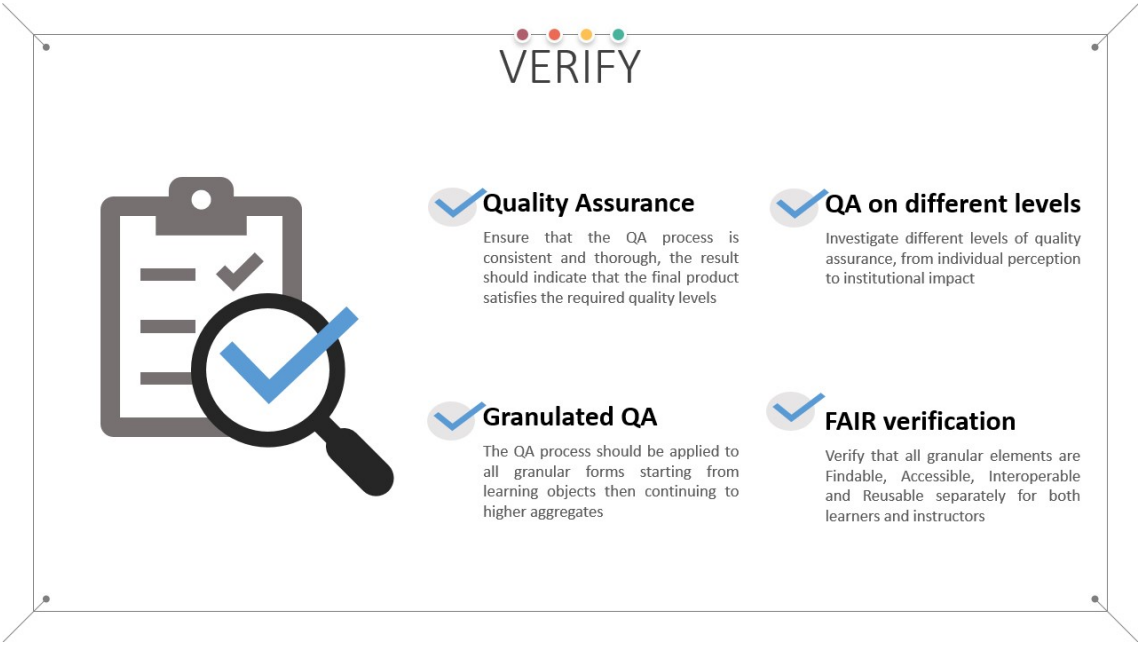

Fig. 13 - Main goals of the verify stage

Quality Assurance checking of the produced learning materials should be performed consistently and thoroughly to ensure that the final products satisfy the required quality levels. After the internal QA check that was aimed at ensuring that all necessary materials are consumable by the target audience, upon publishing it is necessary to ensure the full QA spectrum for the learning materials. In essence this means double checking everything that was performed in the internal QA check but now in a production environment assuming the role of a leaner/instructor/interested party.

During the verification stage various aspects of quality assurance for the learning materials should be analysed. One of the frequently used approach for this goal is the Kirkpatrick model [R95], consisting of the following levels:

- Level 1: Participant Reaction;
- Level 2: Learning;
- Level 3: Job Impact;
- Level 4: Business Impact.

In this model, the first 2 levels impact the learner him/herself, while the latter two levels impact the organization that the learner belongs to.

Specific quality assurance activities can be performed on the learning objects as well as on the higher levels of granularity [R96]. Example traditional QA checklists can be found in [R96] and [R97]. Additional information on quality assurance can also be found in the outputs from Task 2.4 in the Skills4EOSC project.

The FAIR aspects that also need to be verified at this point include:

- Findability - can the content be found on the platform/repository, using a catalogue, or even better, a general-purpose search engine;
- Accessibility - can the learner/instructor access all descriptive details related to the learning materials, are access rules clearly stated, are they correctly implemented;
- Interoperability - standardised metadata description is used for the learning materials based on standardised vocabularies; easy to use, widely available tools are needed to consume the content; standardised editable formats are provided for other instructors;
- Reusability - learners can share the content (under permissive licence) with other peers; other instructors can reuse (as a whole or in part) the content together with a comprehensive facilitators guide while developing their own learning materials under the rules defined with permissive licence (guidance on attribution is provided).

The analysis of already gathered feedback can also help at this stage to verify the QA levels and if all FAIR aspects are implemented as expected.

4.1.7 Continuous improvement

To keep up with the ever-changing environment, as well as the technological changes, continual improvement is considered to be one of the crucial phases in the learning materials development process. The implementation of continuous improvement increases the possibility to deliver successful training and produce adequate, accurate, up-to-date and high-quality learning materials.

Fig. 14 - Continual Improvement Implementation

In many cases, the continual improvement process is an iterative process, with small and measurable steps and outcomes.

The model presented in Fig. 14 can be adapted to the needs of the FAIR-by-design methodology. The basic driver for the process should be the feedback gathered from the learners, as well as the usage of the learning objects by the community of other instructors. The gathered feedback should be continuously analysed and actions for improvement should be taken based on this analysis.

In addition to traditional feedback analysis, the co-creation of learning materials can also be considered as a form of continual improvement. In this model, the learners are considered as partners in the creation and upgrading of the learning materials. Different methodologies exist to employ this model, as presented in [R98].

Regardless of the way the feedback is analysed, the main goal of the continuous improvement stage is to identify clear goals on improvement. Once the goals are set, the FAIR-by-design workflow restarts aiming to produce new, updated and improved learning objects that will implement the envisioned goals. This restart will trigger the creation of a new version of learning materials. Historical versions should be kept for tracking purposes, and the versioning information should be clearly stated in the metadata for the improved materials. The feedback analysis at the end of the next cycle

can be used to measure if the goals set for the new version have been successfully reached.

5 Checklists


A FAIR-BY-DESIGN METHODOLOGY FOR THE FAIR-IFICATION OF EXISTING LEARNING MATERIALS
Use this guide to make your already prepared learning materials FAIR by jumping through the different FAIR-by-design stages
Prepare
Discover
Design
Produce
Publish
Verify
FAIR-by-design methodology workflow stages
Skills 4 eosc
Open Science Commons
FAIR skills
Empower with the skills necessary to implement the FAIR principles. Learn about IPR, metadata, PIDs, repos, …
»Prepare
Versioning
Use a versioning system to implement document version control. Start a new version.
»Prepare
Structure
Decide the level of aggregation on which FAIR will be implemented. The more granular the better reuse.
»Design
Reused?
If you already reused some content you must check if it has permissive license and provide attribution
»Design
Schema
Choose a metadata schema to describe the material, such as the RDA minimal metadata schema
»Produce
Vocabularies
Redefine the main keywords: learning objectives, audience, microcredentials, … using the controlled vocabularies related to the schema
»Prepare
Facilitator kit
Develop an instructor kit (how to, tips & tricks, …) that should help facilitate the process of others reusing the learning material
»Design
Tools & Formats
Use interoperable tools that will enable you to save the intermediate (editable) and final content in open file formats
»Produce
Metadata
Describe the material (all FAIR objects) using the rules provided in the chosen schema
»Produce
Enable reuse
Choose a permissive compatible license for your materials and define the attribution that should be provided if someone decides to reuse them
»Publish
Repositories
Republish the new final version of your learning materials & metadata on the learning platform. Choose a repository for the intermediary files to enable reuse
»Publish
PIDs
Use PIDs for all published FAIR objects, both final and intermediary. Ensure metadata longevity and avoid link rots
PID
»Publish
Access rules
Ensure all material is widely accessible by defining open access rules for learning and reuse
»Publish
Cocreation
Enable the possibility to gather feedback and comments from users and peers
»Publish
Catalogue
Improve findability by creating an entry in a selected training catalogue(s) (thematic, project, regional, general, etc.)
»Publish
QA
Check if materials are FAIR together with quality assurance checklists. Analyse the gathered feedback and start improving
»Verify
CONTINUE TO IMPROVE BY DEVELOPING UPDATED VERSIONS
Supporting eosc
Co-funded by the European Union
UK Research and Innovation

6 Conclusion

Educational resources that aim to upskill relevant stakeholders on the topics related to OS and EOSC should follow the same principles that they promote, i.e. should be FAIR themselves. By incorporating the FAIR principles into the learning materials, the training and skills EOSC pillar can be implemented in a sustainable manner ensuring a strong, collaborative training community that can offer high-quality up-to-date learning materials.

To help achieve this goal, this deliverable has defined a formal methodology for developing FAIR-by-design learning materials by extending the well-adopted backwards instructional design process. Using the outlined steps in each of the process stages, one can make FAIR-aware decisions related to the reuse of existing materials, combining licensing, defining attribution, working with various tools and file formats, and making the final product available for both the target audience and the related community of instructors.

This initial version of the FAIR-by-design methodology will be promoted within the Skills4EOSC training community as well as to the wider EOSC community of trainers. The training materials that will be developed to train the community on how to practically implement the methodology will be devised using the proposed FAIR-by-design methodology itself, thus essentially showcasing the implementation of the proposed steps, recommendations and guidelines. This activity combined with the discussions and feedback gathered from other relevant stakeholders while promoting the methodology will be used as valuable feedback to further improve the methodology and its effective practical implementation. Thus, the proposed methodology will be transformed into a live guidelines document that will be continuously improved through the project lifetime.

7 References

No	*Description/Link*
R1	WIPO, What is Intellectual Property?, 2020
R2	The European Open Science Cloud (EOSC) Strategy 2020-2024
R3	Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018, 2016
R4	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, Publications Office, 2021
R5	Upskilling Countries to Engage in EOSC Task Force
R6	EOSC Future Wiki, Training and Skills Work Package Objectives
R7	RDA FAIR4RS WG, FAIR Principles for Research Software (FAIR4RS Principles), Zenodo, 2022
R8	ENVRI-FAIR, D6.2 FAIR training materials catalogue & integration with Common Training Platform, 2021
R9	ELIXIR Training Platform
R10	Barker, Phil, & Whyte, Angus, Harmonizing Metadata for Exchange of FAIR Training Materials (1.1), Zenodo, 2020
R11	Clare, Helen, Cepinskas, Linas, Leenarts, Ellen, & Krzyzanek, Mateusz, EOSC Synergy Online Training Handbook materials. Zenodo, 2022
R12	Garcia L, Batut B, Burke ML, Kuzak M, Psomopoulos F, Arcila R, et al. Ten simple rules for making training materials FAIR. PLoS Comput Biol 16(5): e1007854,2020
R13	UNESCO, OER
R14	TRIPLE project homepage
R15	Learning resource, International Bureau of Education, UNESCO

R16	Hung, I-Chun, and Nian-Shing Chen. "Embodied interactive video lectures for improving learning comprehension and retention." Computers & Education 117, pp: 116-131, 2018.
R17	OECD, Giving Knowledge for Free, The Emergence of Open Educational Resources, 2007
R18	Wiley, The Instructional Use of Learning Objects, 2002
R19	Examples of OER, digital learning objects
R20	Matt Powel, Using Learning Objects as “Knowledge Packages” to Achieve Learning Objectives, Docebo
R21	Balatsoukas P, Morris A, O'Brien A. Learning objects update: Review and critical approach to content aggregation. Journal of Educational Technology & Society, 11(2):119-30, 2008.
R22	IEEE Standard for Learning Object Metadata
R23	SCORM.com
R24	Metadata: schema development & documentation
R25	Getting started with schema.org
R26	schema.org CreativeWork
R27	schema.org Course
R28	schema.org LearningResource
R29	Bioschemas - Bioschemas
R30	An RDF Vocabulary for Open Educational Resources | Open Educational Resource Schema
R31	N. J. Hoebelheinrich et al., ‘Recommendations for a minimal metadata set to aid harmonised discovery of learning resources’, June 2022
R32	Training Catalogue - Minimal Metadata for Learning Resources - EOSC Future Public - Wiki EOSC Future
R33	OpenPlato Training Catalogue
R34	SSH Training Discovery Toolkit
R35	NI4OS-Europe Training Platform

R36	EOSC portal training catalogue
R37	RDA 'Minimal Metadata Set', Google Docs
R38	E. Newbold, G. Kayumbi, A. Whyte, and E. Lazzeri, 'Summary Report: Workshop on Harmonising Training Resource Metadata for EOSC Communities', May 2021
R39	FAIRsharing | FAIR Metrics - Metadata Longevity
R40	Azami HH, Ibrahim R. Development and evaluation of massive open online course (MOOC) as a supplementary learning tool: An initial study. International Journal of Advanced Computer Science and Applications, 10(7), 2019.
R41	UNESCO & CoL, A Basic Guide to Open Educational Resources, 2015
R42	Cognota, Instructional Design: Everything you need to know
R43	Shift, 10 Super Powers of the World's Greatest Instructional Designer
R44	FAIR guiding principles
R45	DARIAH-Campus, Introduction to persistent identifiers
R46	WIPO, Understanding Copyright and Related Rights, 2016
R47	European Commission, Copyright Policies, 2022
R48	Creative Commons Licences
R49	GNU project, FLOSS and FOSS Licences
R50	Creative Commons Licence Chooser Tool
R51	CC Licences
R52	CC Licence Compatibility Chart
R53	CC Recommended Practices for Attribution
R54	Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society
R55	Samantha Martin-Woodgate, Instructional Design Best Practices, 2021

R56	Krathwohl DR. A revision of Bloom's taxonomy: An overview. Theory into practice, 41(4):212-8, 2002.
R57	Whyte, Angus, Green, Dominique, Avanco, Karla, Di Giorgio, Sara, Gingold, Arnaud, Horton, Laurence, Koteska, Bojana, Kyprianou, Katerina, Prnjat, Ognjen, Rauste, Paivi, Schirru, Luca, Sowinski, Claire, Torres Ramos, Gabriela, & Van Leersum, Nida. (2023).
R58	OER Commons
R59	Browse by Academic Subject | GALILEO Open Learning Materials
R60	OpenMichigan - Find Open Educational Resources
R61	Mason OER Metafinder (MOM) Search
R62	MERLOT
R63	OASIS
R64	OERTX
R65	A. Fouilloux et al., 'Galaxy 101 for everyone', Galaxy Training Network
R66	Bioschemas - 1.0 Release - TrainingMaterial Profile
R67	Galaxy Training Network, Galaxy Training: Introduction to Galaxy Analyses
R68	Zenodo - Research. Shared.
R69	Zenodo - OER Communities
R70	FAIRsharing | Databases
R71	EOSC Pillar Training and Support
R72	GoTriple | Search Resources and Users in Social Sciences and Humanities
R73	ELIXIR TeSS (Training eSupport System)
R74	PaN-Training e-Learning
R75	EOSC-Synergy Platform
R76	Digital Learning Toolkit, Designing a Blueprint
R77	Adam Shaw, Modular Course Design, Center for Teaching and Learning, Wiley Education Services, 2017

R78	Remixing CC-Licensed Work
R79	Copyleft licensing
R80	Slava Todavchich, Moqod.com, Understanding open-source and free software licensing, 2019
R81	Pat Michaels, Key Components of a Facilitator Guide, Great Circle of Learning, 2022
R82	Francesca Di Donato, Lottie Provost, Tiziana Lombardo, Michela Vignoli, Stefanie Pohle, Erzsébet Tóth-Czifra, Yin Chen, & Emilie Blotière. TRIPLE Training Toolkit (0.1). Zenodo, 2022.
R83	The Open Science Training Handbook Git Book
R84	The Open Science Training Handbook Source Files
R85	Zettlr Editor
R86	HedgeDoc Demo
R87	Pandoc Document Converter
R88	Aesoph, Lauri M. "Self-publishing guide." 2018
R89	PCC, Instructional Support, Creating Accessible Content
R90	PCC, Accessibility Handbooks
R91	Ellis, Roger. "Quality assurance for university teaching: Issues and approaches." In Handbook of quality assurance for university teaching, pp. 3-18. Routledge, 2018.
R92	Edutechwiki, Learning Object Repository
R93	Persistent Identifiers: What if? (Learning Object included)
R94	Referencing and citing content – GitHub Pages
R95	Edutechwiki, General learning objects repositories
R96	Kirkpatrick Four Level Model of Training Evaluation
R97	NHS Shared Learning, Quality Assurance Checklists for Evaluating Learning Objects and Online Courses, 2009
R98	zipBoard, eLearning Quality Assurance Checklist with use cases and examples

R99	Co-creation in learning and teaching: the case for a whole-class approach in higher education