PATTERN.
Empowering Open and Responsible Research and Innovation

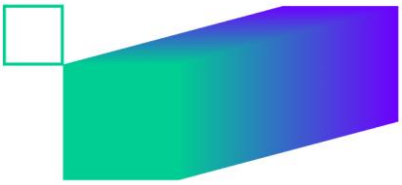
D6.3

First Data Management Plan

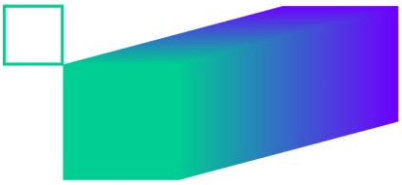
Project Name	Piloting open and responsible Activities and Trainings Towards the Enhancement of Researchers Networks
Project Acronym	PATTERN
Contract No.	101094416
Type of action	HORIZON Coordination and Support Actions
Call	HORIZON-WIDERA-2022-ERA-01: Widening participation and strengthening the European Research Area centres
Topic	HORIZON-WIDERA-2022-ERA-01-44: Developing and piloting training on the practice of open and responsible research and innovation
Start of Project	1 January 2023
Duration	42 months

OUR CONSORTIUM

Funded by the European Union



Deliverable title	First Data Management Plan
Deliverable number	D6.3
Deliverable version	1.0
Previous version(s)	N/A
Contractual date of delivery	30th April 2023
Actual date of delivery	28th April 2023
Nature of deliverable	DMP
Dissemination level	Public
Work Package	WP6
Task(s)	T6.3
Partner responsible	APRE
Author(s)	Claudia Iasillo
DOI	10.5281/zenodo.7885401



Contributors

NAME	ORGANISATION
Alessio Livio Spera	APRE
Cristina Lagido	AU
Pietro Rigonat	LOBA
Asya Salnikova	ESF
Eric Cherel	LPI
Elli Papadopoulou	OpenAIRE
Venkataraman Shanmugasundaram	OpenAIRE
Pascal Flohr	KNAW

Peer Reviews

NAME	ORGANISATION
Alessio Livio Spera	APRE
Jonathan England	OpenAIRE
Marco Soriano	UNISR

Revision History

VERSION	DATE	REVIEWER	MODIFICATIONS
0.1	17/04/2023	Claudia Iasillo (APRE)	First Draft
0.2	25/04/2023	Alessio Livio Spera (APRE), Pietro Rigonat (LOBA), Asya Salnikova (ESF), Eric Cherel (LPI), Jonathan England, Elli Papadopoulou, Venkataraman Shanmugasundaram (OpenAIRE), Marco Soriano (UNISR), Pascal Flohr (KNAW)	Comments from partners
1.0	28/04/2023	Claudia Iasillo	Version submitted

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them

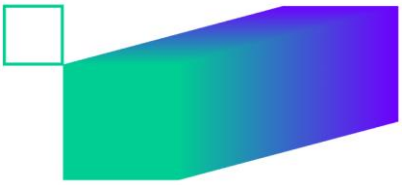
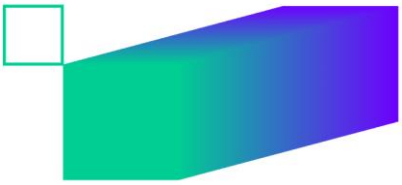


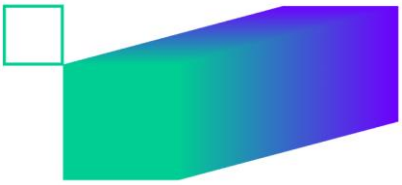
Table of Abbreviations and Acronyms

Abbreviation	Meaning
CS	Citizen Science
D	Deliverable
DMP	Data Management Plan
ERA	European Research Area
FAIR	Findable, Accessible, Interoperable, Accessible
GDPR	General Data Protection Regulation
HEIs	Higher Education Institution
IPR	Intellectual Property Rights
MS	Microsoft
OA	Open Access
OS	Open Science
RRI	Responsible Research and Innovation
WP	Work Package



Index of Contents

Index of Tables

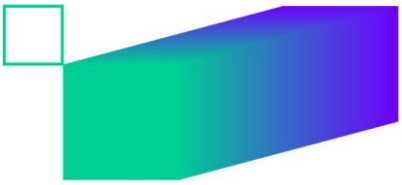


1. Executive Summary

The current document, titled PATTERN First Data Management Plan has been developed within the framework of the PATTERN project which is funded by the European Union's Horizon Europe Research and Innovation Programme under Grant Agreement No 101094416.

This deliverable describes the data management life cycle for the data to be collected, processed and/or generated by the PATTERN project. As part of making research data findable, accessible, interoperable and reusable (FAIR), the Data Management Plan (DMP) provides a summary of the main elements to take into account in the definition of the PATTERN data management policy to be used by project partners throughout the project activities. The DMP will provide indications about the management of all the data generated during the project activities, including how data will be collected, managed, stored and made available during the project and how they will be shared upon PATTERN completion.

The DMP is a living document that will be updated during the project lifetime according to the project's progresses and rising needs. The next versions of PATTERN DMP will describe the procedures applied and it will provide more details about the long-term management and preservation of project data.



2. Introduction

2.1 PATTERN project

PATTERN is a coordination and support action whose aim is to promote the practice of Open and Responsible Research and Innovation (Open RRI) by developing and piloting training activities for researchers at all stages of their careers.

While research institutions and universities have always paid a lot of attention to trainings that build technical and scientific competences of their researchers' workforce, only recently they have better embraced their institutional responsibility in supporting the professional development of researchers beyond academia, by investing resources in offering researchers advanced trainings on more transversal and generic skills (**transferable skills**). PATTERN trainings will strengthen researchers' ORRI transferable skills, with the ultimate goal to empower higher education institutions and research organizations to embrace a transformative process to improve the excellence of the science conducted, the capacity within the European Research Area (ERA) to tackle societal challenges and the interaction between science and society.

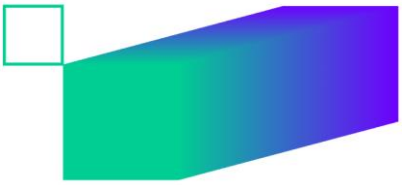
PATTERN will focus on defining a set of training materials and activities on transferable skills within the ORRI framework. Eight main transferable skills have been identified: Open Access (OA); FAIR data management; Citizen Science (CS); Research Integrity; Gender, Non-discrimination and Inclusion in research; Dissemination and Exploitation of results; Science communication; Management and Leadership. For each of them, specific training modules will be developed and tested in 14 Higher Education Institutions (HEIs), namely Pilot Organizations, and made publicly accessible on the PATTERN platform. Moreover, PATTERN will address authorities responsible for researcher training, presenting evidence-based recommendations on educational policies to empower HEIs in contributing to bridge the gap between science and society.

2.2 Objectives and structures of PATTERN DMP

According to the Horizon Europe guidelines, the Data Management Plan (DMP) describes the data management life cycle for the data to be collected, processed and/or generated by the project. As part of making research data findable, accessible, interoperable and re-usable (FAIR), a DMP should include information on:

- the handling of research data during & after the end of the project
- what data will be collected, processed and/or generated
- which methodology & standards will be applied
- whether data will be shared/made open access and
- how data will be curated & preserved (including after the end of the project).

The data accessibility and the long-term data preservation are the ultimate goals of an effective data management, which must be thoroughly planned since the early stages of the project. This deliverable aims at describing the overall data management process within PATTERN: the types of data collected or generated during the project,



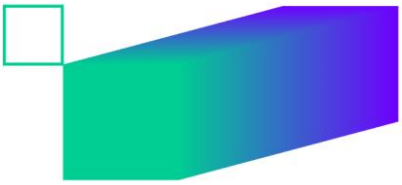
how data will be collected, processed, stored and made available during the project lifetime, and how they will be stored and shared upon PATTERN completion.

The management of the data produced during the project will aim to ensure open access taking also into consideration the “as open as possible, as closed as necessary” principle, to ensure the adoption of suitable measures to preserve project results that can be further commercially exploited in the future (e.g. through patenting). Thus, the project will seek a balance between openness and protection of information, commercialization and Intellectual Property Rights (IPR), privacy concerns, security etc. In case project partners need to keep specific parts of their research data closed, the choice will be explicitly justified in the DMP.

The DMP aims to monitor privacy and confidentiality requirements regarding the generated data. In addition, the DMP aims to ensure that the legal and ethical standards for data generation, use, storage, and sharing are applied throughout the project and in line with the overall management of the project, as foreseen in the Grant Agreement and Consortium Agreement, and as detailed in D6.2 PATTERN Ethics Requirements.

The deliverable represents the first version of the PATTERN DMP, which should be considered as a living document and which will be regularly updated during the project lifetime. Two updates are planned: D6.4 Second Data Management Plan at M18 (June 2024) and D6.5 Final Data Management Plan at M36 (December 2025).

The current document firstly describes the Data Lifecycle within PATTERN (see Section 3 PATTERN Data Lifecycle) detailing the data collection, processing, sharing and preservation. Section 4 FAIR Data describes how data will be managed within PATTERN to ensure the application of FAIR principles during the whole project lifetime. In Section 5 PATTERN Datasets, the dataset produced in each PATTERN Work Package (WP) are described, including the type of data, aim of data, formats, estimated size and the measures applied to make the datasets FAIR. Section 6 Data protection and ethical aspects will detail the procedures to ensure the privacy of Participants to PATTERN activities.



3. PATTERN Data Lifecycle

In the course of the PATTERN project, a variety of data is produced, collected and processed. Appropriate methods and tools for data collection and analysis have to be selected, and a system for storing, organising, and protecting the data collected during PATTERN lifetime must be developed. This includes creating protocols for data collection, processing, and analysis. Another important aspect of defining data management of a research project is to plan consent for gathering and managing certain personal data, which involve obtaining voluntary consent from individuals to participate in project activities (mainly events and surveys/interviews). Informed consent procedures and template are provided in D6.2 PATTERN Ethics Requirements.

3.1 Data Collection

3.1.1 Creating PATTERN data

Since PATTERN is a Coordination and Support Action, no research data will be collected. However, the following categories of data will be collected in the project:

- **Research Data:** recorded information related to technical/scientific activities (e.g. analysis of existing practices, policies, etc.). The collection may require surveys, interviews, instruments to record, transformation of existing datasets to create new data, etc.
- **Stakeholders Data:** such as Contact Details or any other (personal) data of PATTERN stakeholders. The collection may involve registration forms to event, newsletter subscription, website login, etc. The handling of this data will take into account General Data Protection Regulation (GDPR).
- **Platform Data:** platform metrics are collected (visited pages, login times, function usage, etc.) but individual usage data is not published, not shared and only used for technical purposes. Projects data, user profiles are provided by users on a volontary basis and under their full control.

It is worth to point out that administrative/financial data is not included in the research data definition, and in the PATTERN project it will be considered other data. Moreover, the identified categories can occasionally overlap, and data can belong to more than one category depending on the specifics of the data itself.

3.1.2 Documenting PATTERN data

Research data come in various formats: text, numeric, multimedia, models, software languages, discipline and instrument specific, etc. For each WP, the format of the data produced will be described, with the ultimate goal to facilitate the accessibility and interoperability of the data, giving preference to open and standard formats. In this deliverable, a first estimation of the size of the data collected in each WP will also be provided. This information will be updated in the next version of PATTERN DMP according to the actual size of data collected.

The data collected within PATTERN will be organized taking into consideration the following aspects, as far as applicable.

Version Control

To overcome the challenge of managing and tracking research materials during the course of research, especially in collaborative projects, data will be organized to keep track of different versions of the datasets, through the application of a version control system, e.g sharepoint versioning and file naming convention .

Naming conventions used

With the overall aim to make the data accessible (and ultimately foster reproducible research), PATTERN datasets will be carefully named by choosing file names that are informative and useful for both humans and machines. Following, there are some suggestions about the naming conventions within PATTERN.

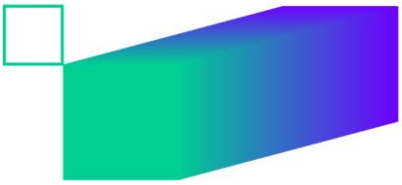
Choose machine readable names
Use deliberate delimits. Common approach is using "_" and "-" to delimit units of metadata in the file names. A general rule is to use "-" to separate words you want to glob together and "_" to separate different information within a file name. Don't use spaces, punctuation, capital letters or special characters (Using $, @, %, #, &,*, (,), !, etc. may have meanings in programming languages).

Choose human readable names
Choose names that explain the content. The more meaningful the name, the more useful it is for human users. The more metadata you store in the name, the less you need to explain elsewhere. Choose short names.

The identifier is a key part of data documentation and metadata. Table 1 shows the codes that can be used for making identifiers.

Table 1 - Examples of files identification

Description	Deliverables	Meetings	Conferences/Events
First Letters	PATTERN	PATTERN	PATTERN
Underscore	_	_	_
Next letters	Deliverable number (Dx.y) [x=WP number, y=deliverable number]	Type of document (i.e. Agenda, Minutes, Presentation). In case of presentations, mention WP number as well.	Event title
Underscore	_	_	_
Next letters	Short explanatory title for the document	Location and date of the meeting separated by underscore	Location and date of the meeting separated by underscore
Underscore		_	_
Next letters		Short name of organisation and initials of presenter	Short name of organisation and initials of presenter
Underscore	_	_	_



Last letters	"v" and number of revision of this specific report v[number].[number] (e.g. v0.1=draft version, v1.0=final version)	“v" and number of revision of this specific report v[number].[number] (e.g. v0.1=draft version, v1.0=final version)]	“v" and number of revision of this specific report v[number].[number] (e.g. v0.1=draft version, v1.0=final version)

Metadata

Data needs to be documented to help interested users to clearly understand and reuse them. For this reason, descriptive and substantive (i.e. how the data should be read or interpreted) metadata will be elaborated and described in a readme.txt file complementing each dataset (see Table 2). For the metadata to be machine-readable it should also be structured (a standard should be followed[1]).

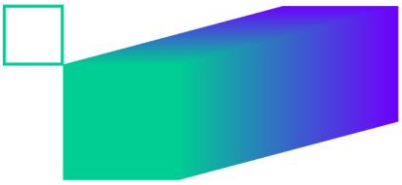
Table 2 – Example of metadata included in the readme file of PATTERN datasets

Creator	Main researchers involved in producing the data
Title	Name or title by which the dataset is known
Contributors	Institutions, organizations or persons responsible for making contributions to the dataset
Publication year	The year when the data was or will be made publicly available
Date of creation	When the resource itself was put together; this could be a date range or a single date
Description	Concise description of the contents of the dataset. Describe the research objective, type of research, method of data collection and type of data
Keywords	Keywords, or key phrases, describing the resource
Spatial coverage	Describe the geographic area to which the data refer (e.g. municipality, town/city, region, country, international)
Persistent Identifier	PIDs will typically be automatically assigned by trustworthy repositories (e.g. Zenodo automatically assigns a DOI). In some cases, a dataset may be known by one or more other (persistent) identifiers
Language	The primary language of the resource
Link to publication	Include the web addresses where the resource is available

3.2 Data processing

Data processing means transforming data into useful information. All data must be checked, cleaned and validated which means that data is correct and relevant. During the data analysis process, data analysis tools are applied in order to transform the data

[1] https://www.dublincore.org/specifications/dublin-core/dcmi-terms/



to easily interpret, share and publish it. The goal of data analysis is to discover useful information in order to draw informing conclusions, and support decision-making.

This DMP outlines how each identified dataset in the PATTERN WPs is being processed (see Section 5 PATTERN Datasets)Administrative project data concerning management and finances is confidential within the consortium and not meant for re-use by third parties. All personal data collected in the course of the project will be kept confidential..

3.3 Data storage

PATTERN involves the collaboration between 19 organisations (i.e. 14 Beneficiaries and 5 Affiliated Entities) and a moderate amount of data will be generated in activities presented in Section 5 PATTERN Datasets. To manage such quantities of data, with the aim to follow the FAIR principles and to stimulate the collaboration among partners, a few data storage tools have been selected. The selection was made having in mind the purposes and the end-users of the data stored, divided in tools appropriate for internal use and daily project activities, and tools for guaranteeing the long-term storage of PATTERN data.

List of data sharing / short-term storage tools NOT appropriate for long-term storing of data generated by PATTERN project but used to carry out daily project activities:

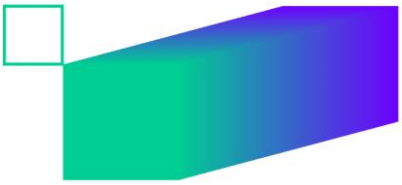
- **Microsoft SharePoint:** is a file hosting service and synchronization service operated by Microsoft as part of its web version of Office. It is a private space, created for the project activities (as detailed in PATTERN D6.1 Project Management Plan), used by all beneficiaries to share working files and store temporary master copy of raw data that should be available for use by other Beneficiaries for data processing.
- **Local drives, company cloud storage and external portable storage devices:** these are storage facilities that do not fall under surveillance of this data management plan.

List of storage tools appropriate for long-term storing of data and to guarantee the findability, accessibility, interoperability, and reusability of data generated by PATTERN:

- **Zenodo**: A PATTERN Zenodo community has been created. PATTERN datasets will be stored there. Data files are versioned. The uploaded data is archived as a Submission Information Package. Derivatives of data files are generated, but original content is never modified. All data files are stored in CERN Data Centres, primarily Geneva, with replicas in Budapest. Data files are kept in multiple replicas in a distributed file system, which is backed up to tape on a nightly basis. All data files are stored along with a MD5 checksum of the file content. Files are regularly checked against their checksums to assure that file content remains constant. Each beneficiary has the possibility to upload the data produced by their project activities in the community and a moderator will validate it. Each data-set will have a readme file.

Link: https://zenodo.org/communities/pattern/

3.4 Data Preservation



All relevant obtained data will be archived in ZENODO. All data and items on ZENODO will be retained for the lifetime of the repository. This is currently the lifetime of the host laboratory CERN, which currently has an experimental programme defined for the next 20 years at least. By sharing the data on the platform (see section above), PATTERN addresses the long-term preservation of the project published data.

3.5 Data Reuse

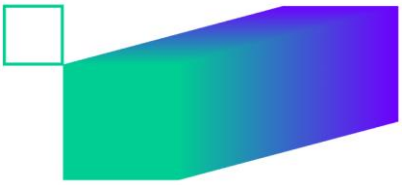
The PATTERN project builds upon existing data and also reuses data. Data collected and produced is made available for reuse whenever it is possible (following the approach as open as possible as closed as necessary). Details concerning the ownership, transfer and dissemination of projects results are defined in section 8 of the PATTERN Consortium Agreement and shall be followed accordingly. The relevant rules of the Grant Agreement, in specific Article 13, Article 15, Article 16 and Annex 5, are also relevant and apply accordingly.

3.5.1 Access to PATTERN Data

PATTERN Beneficiaries will be the main users of data produced in the project. In PATTERN, all data may be classified into open, sensitive and closed. Access depends on the classification of the data.

- **Open data** must in PATTERN be acknowledged by citing the data set. Permission from the researcher is not required.
- **Sensitive (or confidential/restricted) data** may be made available by the researcher after any identifying information has been removed. Then, they can be used and cited. Permission from the researcher could be sought and access can be restricted to certain conditions (e.g. ethical clearance).
- **Closed data** are not available for sharing.

In any case, PATTERN data will have the the latest Creative Commons Attribution licence (currently CC-BY 4.0 International).



4. FAIR Data

The FAIR principles describe how research outputs should be organised so they can be more easily accessed, understood, exchanged and reused. Major funding bodies, including the European Commission, promote FAIR data to maximise the integrity and impact of their research investment[2].

4.1 Findable

Findability of data will be fulfilled by outlining the reference to standard identification mechanisms, the naming conventions used, the approach towards search keywords and clear versioning, as well as specification of standards for rich metadata creation (see section 3.1.2 Documenting PATTERN data).

4.2 Accessible

All PATTERN public deliverables and reports will be openly published on the project website, and on the Zenodo community. All personal data related to the stakeholders are restricted to consortium members. Only anonymised data will be made available to the public (in publications, events etc).

4.3 Interoperable

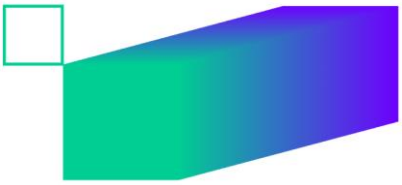
The interoperability of data will be established by specifying what data and metadata vocabularies, standards or methodologies are used.

4.4 Reusable

The increase of data reuse should be documented by specifying how the data are licensed to permit the widest reuse possible. It will also be notified when and to whom the data will be made available for re-use, also after the end of the project, as well as how long the data will remain re-usable (see Section 3.5 Data Reuse).

Administrative project data concerning management and finances is confidential within the consortium and not meant for re-use by third parties. All personal data collected in the course of the project will be kept confidential.

[2] https://www.openaire.eu/how-to-make-your-data-fair

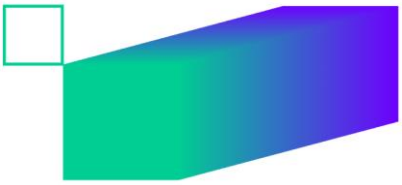


5. PATTERN Datasets

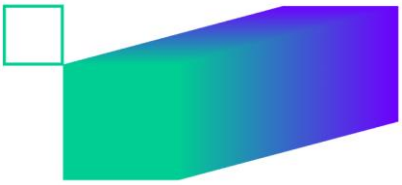
The following sections represent the lifecycle of each dataset produced and processed in the PATTERN project following the FAIR principles.

5.1 Work Package 1

DATA SUMMARY	
DATA GENERATED/COLLECTED	Research Data (e.g. mapping of training activities; survey data entered by users; Interviews data) Stakeholders data (Mutual Learning Events, such as name, affiliation, email)
DATA FORMAT	.XLM .DOC /.DOCX .XLS / .XLSX .PPT / .PPTX .PDF Photo Format (e.g., .JPEG, .PNG) Video Format (e.g., .MKV, .MP4, .AVI)
SIZE OF DATA	1 GB or less
SOFTWARE TOOLS FOR CREATING/PROCESSING/VISUALIZING DATA	MSOffice 365(Excel, Word, PowerPoint, Outlook, MS Teams, MS Sharepoint); Zoom; SPSS, SurveyXact
USE OF PREEXISTING DATA – DATA SOURCE	Yes, training collections on Zenodo. Existing reports, web-pages, and openly available information.
STORAGE AND BACKUP STRATEGY	Microsoft OneDrive, Sharepoint
AUDIENCE FOR REUSE	Researchers, anyone that accesses Zenodo
MAKING DATA FINDABLE	
DATA IDENTIFIERS	DOI for the data uploaded to Zenodo
DIRECTORY AND FILE NAMING CONVENTION	Project and Standard good practices
SEARCH KEYWORDS FOR RE-USE	As per project
VERSIONING MECHANISM	Standard as per project
METADATA STANDARDS AND AUTOMATIC CREATION OF METADATA	Yes. Metadata standards: RDA Minimal Metadata for Learning Resources (with input from SSHOC and EOSC Future, and somewhat adapted). Controlled vocabularies used from EOSC Future,



	SSH Open Cloud, and EU vocabularies. JSON Schema (Zenodo)
MAKING DATA OPENLY ACCESSIBLE	
DATA ACCESS (OPEN, RESTRICTED, CLOSED)	Mapping data: Open after deliverable submission; Survey data: primary data restricted, processed anonymous data open Interview data: closed unless interviewee grants permission at a later date Stakeholder data: restricted to consortium Task organisers and consortium contributors in accordance with GDPR
HOW WILL DATA BE MADE ACCESSIBLE	Zenodo
METHODS OR SOFTWARE TOOLS AND SOFTWARE DOCUMENTATION NEEDED TO ACCESS THE DATA	Web browser, MsOffice 365 (Excel), PDF viewer
DEPOSITION OF DATA AND ASSOCIATED METADATA, DOCUMENTATION AND CODE	No
RESTRICTIONS ON USE AND ACCESS	Stakeholder data: restricted to consortium Task organisers and consortium contributors in accordance with GDPR Interviewee data closed unless interviewee grants permission at a later date.
INFORMATION RETAINED IN ORDER FOR THE DATA TO BE READ AND INTERPRETED IN THE PROJECT AND IN THE FUTURE	Primary data from Survey and interviews will be stored for 5 years on Microsoft OneDrive and then deleted. Mapping data will be integrated in a project deliverable and made available on Zenodo. Stakeholder data kept for duration of project activities then deleted.
MAKING DATA INTEROPERABLE	
DATA AND METADATA VOCABULARIES, STANDARDS OR METHODOLOGIES TO FACILITATE INTEROPERABILITY	Controlled vocabularies used from EOSC Future, SSH Open Cloud, and EU vocabularies.



STANDARD VOCABULARIES FOR ALL DATA TYPES	No
MAPPINGS TO MORE COMMONLY USED ONTOLOGIES	Ontologies as above (RDA Minimal Metadata for Learning Resources[3], EOSC Training Resource Profile[4]; SSHOC has mapped most of these to schema.org[5]). Other ontologies to be chosen at a later stage
MAKING DATA RE-USABLE	
DATA SHARING AND RE-USE	Training mapping data and open access reports and publications may be reused
AUDIENCE FOR REUSE	Researchers, public
PERIOD OF RE-USABILITY	N/A
RESTRICTIONS ON DATA RE-USE (LICENSES)	No
DATA SECURITY, ARCHIVING AND PRESERVATION	
SECURITY MEASURES	standard practice
ARCHIVING OF DATA AND LONG-TERM PRESERVATION	Zenodo

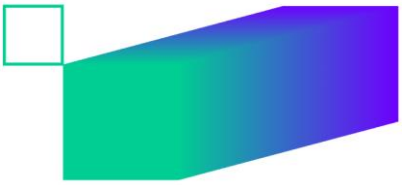
5.2 Work Package 2

DATA SUMMARY	
DATA GENERATED/COLLECTED	Platform Data
DATA FORMAT	Text Data exportable in JSON format
SIZE OF DATA	30 GB

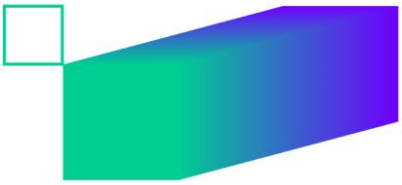
[3] Hoebelheinrich, N. J., Biernacka, K., Brazas, M., Castro, L. J., Fiore, N., Hellström, M., Lazzeri, E., Leenarts, E., Martinez Lavanchy, P. M., Newbold, E., Nurnberger, A., Plomp, E., Vaira, L., van Gelder, C. W. G., & Whyte, A. (2022). Recommendations for a minimal metadata set to aid harmonised discovery of learning resources (Version 1.0). Research Data Alliance. https://doi.org/10.15497/RDA00073

[4] https://wiki.eoscfuture.eu/display/PUBLIC/EOSC+Training+Resource+Profile+-+Data+Model

[5] Research Data Alliance. https://doi.org/10.15497/RDA00073 ; Ricarda Braukmann, Ellen Leenarts, Simon Saldner, Veronika Keck, Judith Wehmeyer, Alejandra Albuerne, Klaus Illmayer, Matej Durco, Vasso Kalaitzi, Tatsiana Yankelevich, Darja Fiser, Ana Cvek, & Anca Vlad. (2022). D6.11 SSHOC Trainer Toolkit (final) (v1.0). Zenodo. https://doi.org/10.5281/zenodo.6564283



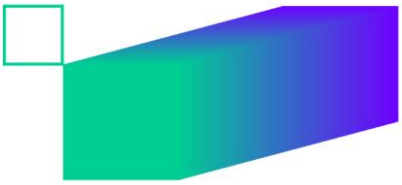
SOFTWARE TOOLS FOR CREATING/PROCESSING/VISUALIZING DATA	VueJS on frontend web side and Python/SQL/Django on backend side
USE OF PREEXISTING DATA – DATA SOURCE	No. Data is exclusively entered voluntarily by users after project start
STORAGE AND BACKUP STRATEGY	Data is stored on Microsoft Azure, with multi-site backup strategy
AUDIENCE FOR REUSE	Researchers
MAKING DATA FINDABLE	
DATA IDENTIFIERS	Open data is accessible through APIs with unique objects identifiers
DIRECTORY AND FILE NAMING CONVENTION	Standard project good practices
SEARCH KEYWORDS FOR RE-USE	As per project
VERSIONING MECHANISM	Data update data
METADATA STANDARDS AND AUTOMATIC CREATION OF METADATA	Yes, standard web development practices. Some metadata are based on Wikipedia identifiers.
MAKING DATA OPENLY ACCESSIBLE	
DATA ACCESS (OPEN, RESTRICTED, CLOSED)	Open only by explicit user choice.
HOW WILL DATA BE MADE ACCESSIBLE	Data accessible on Web pages and directly through APIs
METHODS OR SOFTWARE TOOLS AND SOFTWARE DOCUMENTATION NEEDED TO ACCESS THE DATA	Standard APIs
DEPOSITION OF DATA AND ASSOCIATED METADATA, DOCUMENTATION AND CODE	Technical doc is shared on a Github repository
RESTRICTIONS ON USE AND ACCESS	Access restricted to specific users.
INFORMATION RETAINED IN ORDER FOR THE DATA TO BE READ AND INTERPRETED IN THE PROJECT AND IN THE FUTURE	Data repositories stay on our Azure Cloud at least for the full project duration.
MAKING DATA INTEROPERABLE	
DATA AND METADATA VOCABULARIES, STANDARDS OR	



METHODOLOGIES TO FACILITATE INTEROPERABILITY	Wikipedia as an ontology (with Wiki identifiers)
STANDARD VOCABULARIES FOR ALL DATA TYPES	no
MAPPINGS TO MORE COMMONLY USED ONTOLOGIES	Other ontologies for skills can be added to the metadata, to be chosen at a later stage
MAKING DATA RE-USABLE	
DATA SHARING AND RE-USE	No, unless formally accepted by users who provided the data who own it (project data), to share good practices after PATTERN project ending
AUDIENCE FOR REUSE	Anyone interested
PERIOD OF RE-USABILITY	5 years
RESTRICTIONS ON DATA RE-USE (LICENSES)	no
DATA SECURITY, ARCHIVING AND PRESERVATION	
SECURITY MEASURES	Standard industry security practices
ARCHIVING OF DATA AND LONG-TERM PRESERVATION	Long term storage on separate cloud instance of monthly database snapshots.

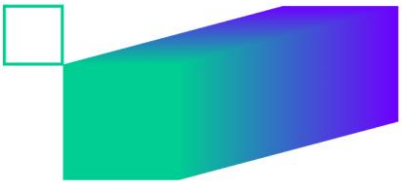
5.3 Work Package 3

DATA SUMMARY	
DATA GENERATED/COLLECTED	Platform data
DATA FORMAT	Text (.DOC /.DOCX, .XLS / .XLSX, .PDF), slides (.PPT / .PPTX), images (e.g., .JPEG, .PNG), video (MP4), audio (MP3), learning resources (SCORM)
SIZE OF DATA	Estimation: >10 kB (text files) to <2GB (video files)
SOFTWARE TOOLS FOR CREATING/PROCESSING/VISUALIZING DATA	Word processors, etc - MS Office (Word, Excel, PowerPoint), learning resource authoring tool - Articulate Rise 360, audio/video editors - TBD



USE OF PREEXISTING DATA – DATA SOURCE	Yes – from other training catalogues, repositories, Zenodo, etc.
STORAGE AND BACKUP STRATEGY	Networked storage, multi-site
AUDIENCE FOR REUSE	Researchers, research support staff, funders, policymakers, citizen scientists, etc.
MAKING DATA FINDABLE	
DATA IDENTIFIERS	PIDs where applicable, typically DOIs
DIRECTORY AND FILE NAMING CONVENTION	Standard project good practices
SEARCH KEYWORDS FOR RE-USE	Training, learning, skills
VERSIONING MECHANISM	As per project
METADATA STANDARDS AND AUTOMATIC CREATION OF METADATA	Yes. RDA ETHRD IG Minimum Metadata Set for Learning Resources
MAKING DATA OPENLY ACCESSIBLE	
DATA ACCESS (OPEN, RESTRICTED, CLOSED)	Open
HOW WILL DATA BE MADE ACCESSIBLE	OpenPlato, Zenodo
METHODS OR SOFTWARE TOOLS AND SOFTWARE DOCUMENTATION NEEDED TO ACCESS THE DATA	Web browser, standard APIs
DEPOSITION OF DATA AND ASSOCIATED METADATA, DOCUMENTATION AND CODE	No
RESTRICTIONS ON USE AND ACCESS	None
INFORMATION RETAINED IN ORDER FOR THE DATA TO BE READ AND INTERPRETED IN THE PROJECT AND IN THE FUTURE	Metadata of all deposited digital objects.
MAKING DATA INTEROPERABLE	
DATA AND METADATA VOCABULARIES, STANDARDS OR	EOSC Future[6]

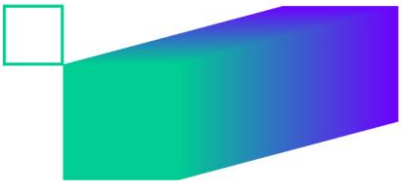
[6] https://wiki.eoscfuture.eu/display/PUBLIC/EOSC+Training+Resource+Profile+-+Data+Model



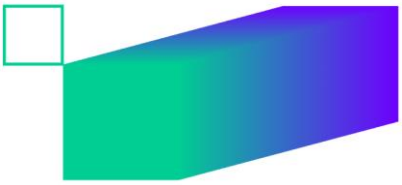
METHODOLOGIES TO FACILITATE INTEROPERABILITY	
STANDARD VOCABULARIES FOR ALL DATA TYPES	No
MAPPINGS TO MORE COMMONLY USED ONTOLOGIES	DCMI
MAKING DATA RE-USABLE	
DATA SHARING AND RE-USE	Yes. CC-BY 4.0 International
AUDIENCE FOR REUSE	All
PERIOD OF RE-USABILITY	Indefinite
RESTRICTIONS ON DATA RE-USE (LICENSES)	Attribution
DATA SECURITY, ARCHIVING AND PRESERVATION	
SECURITY MEASURES	EOSC AAI
ARCHIVING OF DATA AND LONG-TERM PRESERVATION	OpenPlato, Zenodo, cloud (e.g. SharePoint)

5.4 Work Package 4

DATA SUMMARY	
DATA GENERATED/COLLECTED	Research data – grid of mapped EU lvl policies, report of policy analysis, extract from Miro boards/posters from Open Studios Stakeholder data – contact information (name, email, gender, affiliation) of participants to Open Studios
DATA FORMAT	.xlsx, .pdf, .jpeg
SIZE OF DATA	Couple of MB
SOFTWARE TOOLS FOR CREATING/PROCESSING/VISUALIZING DATA	MS Excel, Miro, camera
USE OF PREEXISTING DATA – DATA SOURCE	No



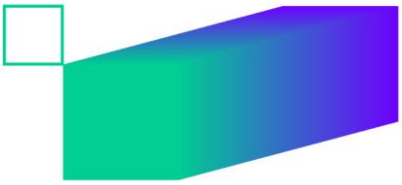
STORAGE AND BACKUP STRATEGY	Institutional intranet, local drives (incl. work OneDrive) and project SharePoint.
AUDIENCE FOR REUSE	Social science researchers
MAKING DATA FINDABLE	
DATA IDENTIFIERS	DOI for the data uploaded to Zenodo
DIRECTORY AND FILE NAMING CONVENTION	According to the project and good practices
SEARCH KEYWORDS FOR RE-USE	As per project
VERSIONING MECHANISM	Standard
METADATA STANDARDS AND AUTOMATIC CREATION OF METADATA	JSON Schema (Zenodo)
MAKING DATA OPENLY ACCESSIBLE	
DATA ACCESS (OPEN, RESTRICTED, CLOSED)	Stakeholder data: closed Research data: open
HOW WILL DATA BE MADE ACCESSIBLE	Zenodo
METHODS OR SOFTWARE TOOLS AND SOFTWARE DOCUMENTATION NEEDED TO ACCESS THE DATA	Internet connection, MS Office (Excel), pdf viewer
DEPOSITION OF DATA AND ASSOCIATED METADATA, DOCUMENTATION AND CODE	no
RESTRICTIONS ON USE AND ACCESS	Stakeholders' data will be restricted and not published, available only for PATTERN beneficiaries and only for the purpose of related project activity (e.g., Open Studio)
INFORMATION RETAINED IN ORDER FOR THE DATA TO BE READ AND INTERPRETED IN THE PROJECT AND IN THE FUTURE	
MAKING DATA INTEROPERABLE	
DATA AND METADATA VOCABULARIES, STANDARDS OR METHODOLOGIES TO FACILITATE INTEROPERABILITY	WIkipedia



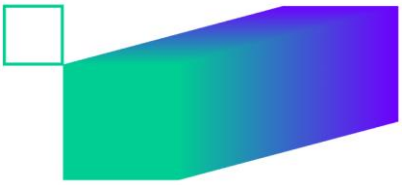
STANDARD VOCABULARIES FOR ALL DATA TYPES	No
MAPPINGS TO MORE COMMONLY USED ONTOLOGIES	no
MAKING DATA RE-USABLE	
DATA SHARING AND RE-USE	Yes, research data may be used for more in-depth policy analysis
AUDIENCE FOR REUSE	Social Science researchers
PERIOD OF RE-USABILITY	n/a
RESTRICTIONS ON DATA RE-USE (LICENSES)	no
DATA SECURITY, ARCHIVING AND PRESERVATION	
SECURITY MEASURES	Standard practice
ARCHIVING OF DATA AND LONG-TERM PRESERVATION	Zenodo for technical data The Stakeholders' data will be purged after maximum of up to 5 years after the end of the project

5.5 Work Package 5

DATA SUMMARY	
DATA GENERATED/COLLECTED	Stakeholders data: Standard data collected by Google analytics, such as: number of users visiting the website, countries of visitors, devices used to visit the website, channels driving the traffic, users' behavior when navigating the website, most visited web pages, etc.
DATA FORMAT	Text Data exportable in CSV format
SIZE OF DATA	N/A
SOFTWARE TOOLS FOR CREATING/PROCESSING/VISUALIZING DATA	Google Analytics
USE OF PREEXISTING DATA – DATA SOURCE	N/A
STORAGE AND BACKUP STRATEGY	Google Analytics (storage), Google Drive (backup)



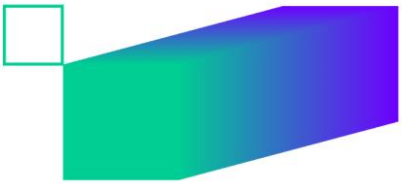
AUDIENCE FOR REUSE	N/A
MAKING DATA FINDABLE	
DATA IDENTIFIERS	Open data is accessible through APIs with unique objects identifiers
DIRECTORY AND FILE NAMING CONVENTION	Standard project good practices
SEARCH KEYWORDS FOR RE-USE	According to project
VERSIONING MECHANISM	Data update data
METADATA STANDARDS AND AUTOMATIC CREATION OF METADATA	Yes, standard web development practices. Some metadata are based on Wikipedia identifiers.
MAKING DATA OPENLY ACCESSIBLE	
DATA ACCESS (OPEN, RESTRICTED, CLOSED)	Open only by explicit user choice.
HOW WILL DATA BE MADE ACCESSIBLE	Data accessible on Web pages and directly through APIs
METHODS OR SOFTWARE TOOLS AND SOFTWARE DOCUMENTATION NEEDED TO ACCESS THE DATA	Standard APIs
DEPOSITION OF DATA AND ASSOCIATED METADATA, DOCUMENTATION AND CODE	Technical doc is shared on a Github repository
RESTRICTIONS ON USE AND ACCESS	Access restricted to specific users.
INFORMATION RETAINED IN ORDER FOR THE DATA TO BE READ AND INTERPRETED IN THE PROJECT AND IN THE FUTURE	N/A
MAKING DATA INTEROPERABLE	
DATA AND METADATA VOCABULARIES, STANDARDS OR METHODOLOGIES TO FACILITATE INTEROPERABILITY	Wikipedia as an ontology (with Wiki identifiers)
STANDARD VOCABULARIES FOR ALL DATA TYPES	No
MAPPINGS TO MORE COMMONLY USED ONTOLOGIES	N/A
MAKING DATA RE-USABLE	



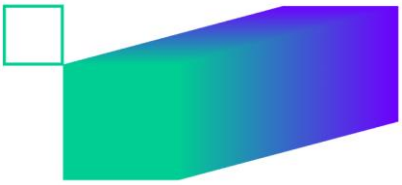
DATA SHARING AND RE-USE	No, unless formally accepted by users who provided the data who own it (project data), to share good practices after PATTERN project ending
AUDIENCE FOR REUSE	Anyone interested
PERIOD OF RE-USABILITY	5 years
RESTRICTIONS ON DATA RE-USE (LICENSES)	No
DATA SECURITY, ARCHIVING AND PRESERVATION	
SECURITY MEASURES	Standard project security practices
ARCHIVING OF DATA AND LONG-TERM PRESERVATION	Long term storage on separate cloud instance of monthly database snapshots.

5.6 Work Package 6

DATA SUMMARY	
DATA GENERATED/COLLECTED	Other, including administrative data (such as deliverables, official documents, meetings recordings and agendas, reports, contact details) and financial data (such as financial reports and statements)
DATA FORMAT	• .DOC /.DOCX • .XLS / .XLSX • .PPT / .PPTX • .CSV • .PDF • Picture Format (e.g., .JPEG, .PNG) • Video Format (e.g., .MKV, .MP4, .AVI)
SIZE OF DATA	As of M4: around 2GB. Estimation: around 15GB by M42.
SOFTWARE TOOLS FOR CREATING/PROCESSING/VISUALIZING DATA	Microsoft Office 365 (including Outlook, MS Teams, MS SharePoint); Adobe Acrobat DC.
USE OF PREEXISTING DATA – DATA SOURCE	No
STORAGE AND BACKUP STRATEGY	A dedicated PATTERN SharePoint site has been created as main tool for storage and backup of data

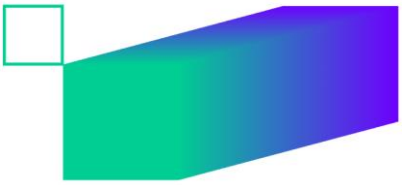


AUDIENCE FOR REUSE	PATTERN Beneficiaries
MAKING DATA FINDABLE	
DATA IDENTIFIERS	• Clear folders and documents naming, single URLs for documents' editing for all Beneficiaries • Standard templates for recurrent documents (agendas, deliverables, etc.)
DIRECTORY AND FILE NAMING CONVENTION	• Each project activities has specific dedicated folders on SharePoint (WPs, Deliverables, meetings, events, etc.), clearly mentioning relevant details such as date, number, versioning, etc. • File naming procedures setup in D6.1 Project Management Plan: i.e., for deliverables, "PATTERN_D[number of deliverable]_[NameOfDeliverable] _v[version number].docx"
SEARCH KEYWORDS FOR RE-USE	N.A.
VERSIONING MECHANISM	• Deliverables: "v0.x" at the end of the file's name for draft versions; "v1.X"at the end of the file's name for final/submitted versions • Generic documents: "YYYYMMDD" at the end of the file's name
METADATA STANDARDS AND AUTOMATIC CREATION OF METADATA	No
MAKING DATA OPENLY ACCESSIBLE	
DATA ACCESS (OPEN, RESTRICTED, CLOSED)	All data will be open for PATTERN Beneficiaries and open, restricted or closed for the general public according to Horizon Europe and Grant Agreement regulations and dissemination levels (e.g. SEN deliverables will not be public)
HOW WILL DATA BE MADE ACCESSIBLE	• Consortium level: all data will be on PATTERN SharePoint site • Public level: public documents will be available on PATTERN website and on the Zenodo community; some of them (e.g. public deliverables) will be available on Cordis



METHODS OR SOFTWARE TOOLS AND SOFTWARE DOCUMENTATION NEEDED TO ACCESS THE DATA	• Web browser, valid email address (to access upon invitation PATTERN SharePoint site), Microsoft Office 365 (including Outlook, MS Teams, MS SharePoint); Adobe Acrobat DC
DEPOSITION OF DATA AND ASSOCIATED METADATA, DOCUMENTATION AND CODE	No
RESTRICTIONS ON USE AND ACCESS	• PATTERN SharePoint site is available for the project's Beneficiaries only • In general, Grant Agreement's restrictions are applied (e.g. deliverables dissemination levels)
INFORMATION RETAINED IN ORDER FOR THE DATA TO BE READ AND INTERPRETED IN THE PROJECT AND IN THE FUTURE	N.A.
MAKING DATA INTEROPERABLE	
DATA AND METADATA VOCABULARIES, STANDARDS OR METHODOLOGIES TO FACILITATE INTEROPERABILITY	EuroVoc, Microsoft Office
STANDARD VOCABULARIES FOR ALL DATA TYPES	N.A.
MAPPINGS TO MORE COMMONLY USED ONTOLOGIES	N.A.
MAKING DATA RE-USABLE	
DATA SHARING AND RE-USE	All data will be stored on SharePoint until 5 years after the end of the project, in compliance with record-keeping indication of the Grant Agreement Data Sheet
AUDIENCE FOR REUSE	PATTERN Beneficiaries, European Commission
PERIOD OF RE-USABILITY	5 years after the end of the project
RESTRICTIONS ON DATA RE-USE (LICENSES)	Only PATTERN Beneficiaries and EC
DATA SECURITY, ARCHIVING AND PRESERVATION	
SECURITY MEASURES	• Microsoft's regular security measures and features

	• The project coordinator (APRE) is in charge of constantly monitor data security, archiving and preservation
ARCHIVING OF DATA AND LONG-TERM PRESERVATION	All data will be stored on SharePoint until 5 years after the end of the project, in compliance with record-keeping indication of the Grant Agreement Data Sheet. Public deliverables will be uploaded on Zenodo as well.



6. Data protection and ethical aspects

This section outlines some key legal and ethical aspects, when it comes to processing of personal data. More details are provided in the D6.2 PATTERN Ethical Requirements.

6.1 Personal Data collection

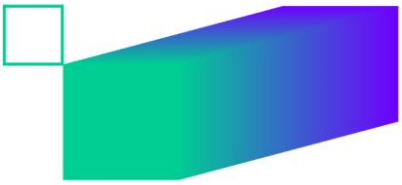
According to Art. 4(1) of GDPR, **personal data** 'means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person'. Examples of personal data protection typically cover: names and surname, home and email address, an identification card number, location data and IP address or a cookie ID. **Processing of personal data** is defined in Art. 4(2) of the Regulation and means 'any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organization, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction'.

Personal data is collected and processed in the following project related activities and tasks:

- Mutual Learning Events
- Workshops, webinars, events, etc.
- Website
- Platform

In the PATTERN project, all data processing complies with EU law as well as national data laws. In its Art. 5, the GDPR provides for the **principles of personal data processing**, stating that personal data must be:

- 'processed lawfully, fairly and in a transparent manner in relation to the data subject ('lawfulness, fairness and transparency');
- collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes (…) ('purpose limitation');
- adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed ('data minimization');
- accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that personal data that are inaccurate, having regard to the purposes for which they are processed, are erased or rectified without delay ('accuracy');
- kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed (…) ('storage limitation');



- processed in a manner that ensures appropriate security of the personal data, including protection against unauthorized or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organizational measures ('integrity and confidentiality').

In order to fulfill to the GDPR requirements, there is a set of indications that each PATTERN Beneficiary dealing with data processing will respect during the project implementation:

1) Consent

Participation of third parties in the PATTERN project is voluntary and individual participants agree to participate in project activities (both online and physical) based on the invitation issued. The consent should be given by an individual (data subject) in a free, specific, informed and unambiguous manner, by way of a request presented in clear and plain language. Moreover, the consent should be given by an affirmative act, such as checking a box online or signing a form. Therefore, along with an invitation, participants receive an informed consent (see D6.2 Ethics Requirements) and agree to participate, before the commencement of any project activity requiring their involvement, in compliance with GDPR.

2) Providing transparent information

Individuals must be clearly informed on who is processing their personal data as well as what data will be processed, why and how. In addition to the above information, individuals should be also informed about who will receive the data, how long the data will be stored, the individual's data protection rights (i.e. right to access, correction, erasure, restriction, objection, portability, etc.), whether there is a statutory or contractual obligation to provide the data and how consent can be withdrawn.

3) Ensuring the right to access and right to data portability

Individuals must be ensured with the right to request access to their personal data, free of charge and in an accessible format. In addition, when the processing is based on consent or a contract, the individual can ask for their personal data to be returned or transmitted to another company (right to portability).

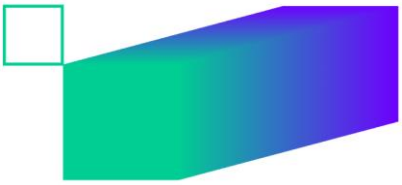
4) Ensuring the right to erasure (right to be forgotten)

An individual has the right to request the data controller to erase their personal data, such as when the data is no longer needed to fulfil the processing purpose. The data controller is not obliged to comply with such request if: the processing is necessary to respect one's freedom of expression and information, they must keep the personal data to comply with a legal obligation; there are other reasons of public interest to keep the personal data, such as public health or scientific and historical research purposes; they need to keep the personal data to establish a legal claim.

5) Ensuring the right to correct and right to object

If an individual thinks that their personal data is incorrect, incomplete or inaccurate, they have the right to have it rectified or completed without undue delay.

6) Obligation to appoint a Data Protection Officer



An entity must appoint a DPO when: it regularly or systematically monitors individuals or process special categories of data; this processing is a core business activity; and it does it on a large scale. In the case of PATTERN Beneficiaries, all relevant declarations and the list of DPOs with their contact details are described in D6.2 PATTERN Ethics Requirements/uploaded on PATTERN SharePoint Site.

7) Obligation of data protection by design and by default

In order to minimize privacy risks and increase trust, a data controller must take all necessary steps (e.g. pseudonymization) to implement the data protection principles and protect the rights of individuals (data protection by design). In accordance with the obligation of data protection by default, the entity processing personal data should ensure that the most privacy friendly setting is the default setting.

8) Obligation to provide proper notification in the case of a data breach

In case a data breach occurs, and the breach poses a risk to individual rights and freedoms, the entity processing personal data should notify the relevant Data Protection Authority within 72 hours after becoming aware of the breach. If the data breach poses a high risk to those affected, the entity may also be required to inform all individuals affected by the data breach.

PATTERN.

Empowering Open and Responsible Research and Innovation

OUR CONSORTIUM

APRE – Agenzia per la Promozione della Ricerca Europea
AARHUS UNIVERSITY
LOBA®
SCIENCE CONNECT – YOUR PARTNER IN SCIENCE
ZSI
OpenAIRE
Open AIRE affiliated entities
SISSA
PLANET INSTITUTE
EARMA – European Association of Research Managers and Administrators
UniSR – Università Vita-Salute San Raffaele
DANS – Data Archiving and Networked Services
UNIVERSITY OF HELSINKI
Trinity College Dublin – Coláiste na Tríonóide, Baile Átha Cliath – The University of Dublin
UNIVERSITY of DEBRECEN
IZMIR INSTITUTE OF TECHNOLOGY
RB
SciLink
Universidade do Minho
HEALLINK

Funded by the European Union

pattern-openresearch.eu info@pattern-openresearch.eu