

DELIVERABLE REPORT

Grant Agreement number:	693989
Project₁ Acronym:	ENLIVEN
Project title:	Encouraging Lifelong Learning for an Inclusive and Vibrant Europe
Start date of the project:	01/10/2016
Duration of the project:	36 months
Deliverable No.	11.1
Deliverable Title:	Data Management Plan
Dissemination level:	Public
Date of submission:	28/9/17
Version:	2 (Month 12)



Encouraging Lifelong Learning
for an Inclusive & Vibrant Europe

Data Management Plan

(Deliverable 11.1)

September 2017

Table of Contents

1	Introduction	6
1.1	Project Summary.....	6
1.2	Types of Data	6
1.3	Organisation of the ENLIVEN project.....	7
2	Data Management.....	8
2.1	Principles	8
2.2	Organisation	9
2.3	Types of Data	9
2.3.a	Aggregated secondary statistical data.....	9
2.3.b	Secondary data in form of anonymised micro data sets:	9
2.3.c	Primary data collection	9
2.4	Data Use and Protection.....	10
2.4.a	Secondary analysis of EUROSTAT and related scientific-use micro data files	10
2.4.b	Interview Transcription and Data Protection	11
2.5	Documentation	12
2.5.a	Documentation Reports.....	12
2.5.b	Project Management Documents.....	13
2.5.c	Website	13
2.6	Hardware and Software.....	13
2.6.a	Intelligent Decision Support System	13
2.6.b	Data Backup and Recovery	13
2.7	Intellectual Property and Ownership	13
2.7.a	Intellectual Property	13
2.7.b	Joint ownership.....	14
2.8	Open Access.....	14
2.9	Quality Assurance	14
3	Work-Package-specific Data Issues.....	15
3.1	WP1: Mapping European and national policies and programmes, and their contribution to economic and social inclusion	15
3.2	WP2: Constraints and facilitators of access and participation	15
3.3	WP3: The role of European governance in adult education & learning policy	15
3.4	WP4: Improving our understanding of the effect of system characteristics by building stronger data and adding a longitudinal, regional & sectoral focus	16
3.5	WPs5–7: Studying the role of workplace learning and patterns of work organisations for early career structuration; qualitative interviews on learning biographies.....	16
3.6	WP8: Knowledge discovery on evidence-based policy making in participating countries; & WP9: Establishment of Intelligent Decision Support System for evidence-based policy making	16
3.7	WP10-11: Dissemination and Project Management & Integration.....	17
4	Appendices.....	18

4.1 Table 1: Key elements of the framework to ensure anonymization within the ENLIVEN research process (for storage/use within the project): (to be refined within the research project) 18

4.2 Table 2: Processing of data in the qualitative research implemented by the ENLIVEN project..... 19

Revisions made to Data Management Plan (D11.1)

This document is to summarise any revisions made to the Data Management Plan throughout the course of the project

Data Management Plan Version	Section	Content	Changes made to content	Version changes made to
Version 1- submitted March 2017	2.5.b	All internal board, committee, work package and meeting/workshop reports (unless specifically deemed confidential by the Co-ordinator or the team members concerned) will be included in the ENLIVEN Dropbox folder.	All internal board, committee, work package and meeting/workshop reports will not be included in the ENLIVEN Dropbox folder, but will be stored on the University of Nottingham's system.	Version 2- September 2017 (Month 12 report)
	2.5.b	On the basis of this management documentation, bi-annual reports on the project's activities and progress will be prepared. These reports will be approved by the Management Board. The MB will also approve a public version that excludes any sensitive or IPR-restricted information; this will be made available on the project's website.	On the basis of this management documentation, annual reports on the project's activities and progress will be prepared (using the quarterly Coordinators Reports used for Management Board Meetings). These reports will be approved by the Management Board. One report will be prepared which will be for public use (e.g. on website and for Advisory Board). However, if there is any sensitive or IPR-restricted information there will be a 'private' version too that this can be included in.	
	2.8	Where Green route publishing is not possible, authors will be provided with funding to publish through the Gold route.	Where Green route publishing is not possible, subject to resources authors will be provided with funding to publish through the Gold route.	

1 Introduction

1.1 Project Summary

ENLIVEN ('Encouraging Lifelong Learning for an Inclusive and Vibrant Europe' – H2020 Young Society 2015) is a research project supported by the European Commission's Horizon 2020 research framework (Project No. 693989). The project responds to Call YOUNG-3-2015 ('Lifelong learning for young adults: better policies for growth and inclusion in Europe'). Its duration is 36 months and it commenced on 1st October 2016.

ENLIVEN's overarching objective is to provide an innovative model and mechanism to support policy debate, policy formation and policy evaluation in lifelong learning, focussing on the needs of today's young adults, and integrating theoretical and empirical perspectives from social and computer sciences. It will generate an evidence-based analysis of where, when and why policies have been effective, and develop a computer-based intelligent system to improve policy-making. The project draws on two research fields – social science and computer science – and combines expertise from both. Partner research institutions are located in nine countries (Australia, Austria, Belgium, Bulgaria, Estonia, Italy, Slovakia, Spain, and United Kingdom); some research tasks will also be undertaken under a specific arrangement with researchers at a university in Denmark.

Objectives

The specific objectives of ENLIVEN are to:

- Map and critically assess key elements of programmes implemented at EU, national and regional levels to support access to and participation in adult learning among excluded population groups and those at risk of social exclusion, assess how these have addressed disadvantage, inequality, and social exclusion, and helped overcome barriers to participation, and in what ways participation in education and training benefits the social and economic inclusion of population groups suffering from exclusion and cumulative disadvantage.
- Assess the impact of "system characteristics" (of initial and adult education, the labour market, the economy, and social protection) on aggregate participation rates (overall, and in various segments of adult education markets), and on the distribution of participation (with special reference to disadvantaged young adults and using gender-sensitive approaches).
- Assess the role of lifelong learning in developing a productive, efficient and competitive economy through investigating what learning potential and innovation ability exists within workplaces, what organisational models favour innovation and innovative training, and how effective learning actions are.
- Identify and map the nature and availability of data about adult and lifelong learning, and integrate these with new research findings from across the ENLIVEN project and, using data mining, establish a knowledge base for the development of an Intelligent Decision Support System to support policy making;
- Design and implement an Intelligent Decision-making Support System (IDSS), and test how this could adapt to new knowledge and learn from restoring and updating users' experience interactively.

1.2 Types of Data

The main types of data the ENLIVEN project will be handling are as follows:

- *Existing national and international datasets.* These are both quantitative (e.g. LFS, AES, PIAAC, ESS) and qualitative (e.g. reports from Eurydice, Cedefop). All these data are either anonymised (in the case of the quantitative data), or in the public domain.¹ None contain private or personal details regarding identifiable individuals. We will ensure we utilise existing datasets wherever possible, to ensure best practice in the project.
 - Existing collections of aggregated data (e.g. Eurostat’s dissemination database, UNESCO’s database on education).
 - Existing accessible quantitative scientific-use microdata sets e.g. LFS, AES, CVTS obtained from Eurostat
 - Other quantitative scientific-use micro datasets not obtained from Eurostat (e.g. PIAAC data obtained from the OECD)
 - Reports from (mainly) European agencies, containing aggregated statistical data and various forms of qualitative data e.g. Eurydice, Cedefop

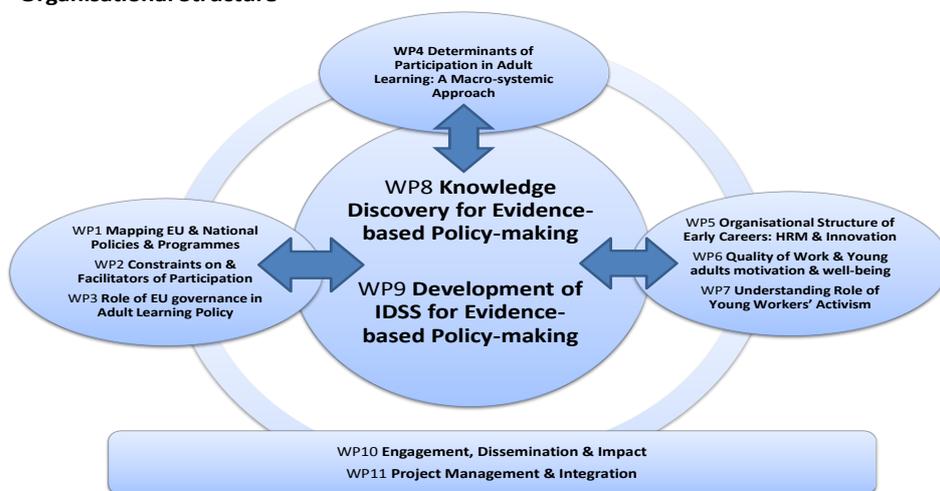
- *Fieldwork data generated by ENLIVEN researchers.* This will come from interviews with policy actors, programme managers, and young adults in educational, community and workplace situations. Fieldwork research will be conducted in a sample of EU member states selected to represent a diversity of socio-economic characteristics and institutional environments (AT, BE, BG, DK, EE, ES, IT, SK, UK), and in Australia.
 - Qualitative Interviews with participants in educational programmes and with participants in organisational case studies (managers; workers in their first 10 years of employment) and their analysis
 - Data collection within the participatory observation/action research phase of the case studies (in WP5) and their analysis.
 - Expert Interviews; Interviews with representatives of educational providers/policy makers in the field of adult learning (WPs 1, 2, and 3) and with business interest organisations and trade unions/social movement organisations (WP7) and their analysis

1.3 Organisation of the ENLIVEN project

The ENLIVEN project is organised in eleven work packages (WPs). Nine of these involve research of various kinds; one is concerned with dissemination of the research and engagement with policy and practitioner communities; one is concerned with project management and the integration of the various elements of the research. The research WPs can be considered in three clusters: each involves distinct types of empirical research. The broad relationship between these WPs is shown in Fig. 1.1.

¹ In a few cases, we may be given access to confidential reports, e.g. generated by or for the European Commission. Such material will only be obtained with appropriate authorisation from the agency concerned, and be treated in accordance with any requirements which the agency imposes.

**Figure 1.1: ENLIVEN:
Organisational Structure**



2 Data Management

2.1 Principles

Data Management Plans (DMPs) are a key element of good data management. The DMP for Enliven has been written with reference to the Guidelines on Data Management in Horizon 2020 which specifically ask for provisions to be made 'FAIR', that is findable, accessible, interoperable and reusable. ENLIVEN will strictly adhere to the 'Data Protection Directive' (Directive 95/46/EC) 16 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, the Charter of Fundamental Rights of The European Union and the European Charter for Researchers, including the Code of Conduct for the Recruitment of Researchers. The European Commission has now issued a new directive on the Protection of Personal Data. The Regulation entered into force on 24 May 2016, and it will apply from 25 May 2018. The ENLIVEN researchers will comply with its requirements.

The DMP covers:

- What data the project will produce
- How the data will be used, managed and stored
- How the data will be made accessible for future research

Additional information on data preservation and data sharing protocols are covered in full in our *Deliverable 11.2 Ethical Procedures (Part A)*.

The Project Team will seek guidance from the University of Nottingham's relevant Professional Services where any uncertainties occur, including Libraries, Research and Learning Resources (LRLR) and Information Services (IS).

2.2 Organisation

The Coordinator, Professor John Holford, and the ENLIVEN Management Board, supported by the Project Co-ordination Team at the University of Nottingham (a Senior Research Fellow, Sharon Clancy, and a Project Administrator, Ruth Elmer), are responsible for ensuring that data are handled by all Consortium Members according to the DMP.

The DMP will be updated as required through the life of the project. Responsibility for this lies with the Project Co-ordinator, supported at Nottingham by the Project Administrator and the Senior Research Fellow.

DMP updating will take place whenever significant changes arise. These might include new data emerging, changes in consortium policies, changes in consortium composition and external factors (e.g. new consortium members joining or old members leaving). In any case, the DMP will be reviewed on a six monthly basis at every second Management Board meeting. It will be a standing agenda item for the meetings, to allow for discussion and updating on an ongoing basis.

2.3 Types of Data

Beyond research literature and reports published by various authorities and institutions, the types of data which ENLIVEN will manage include:

- Aggregated secondary statistical data (tables) obtained from national statistical providers, Eurostat or other European or transnational data providers (CEDEFOP; Eurofound; OECD, UNESCO)
- Secondary data in the form of anonymised scientific use micro data sets, obtained mainly from Eurostat, and partly from other European or transnational organisations (e.g. ESS data from the ESS ERIC; the EJS data from CEDEFOP², PIAAC data from the OECD);³ and
- primary data (gathered by project researchers), including data based on qualitative non- or semi-standardised interviews, small-scale written surveys and data gathered via participator observation during the field work (action research phase) in WP 5.

2.3.a Aggregated secondary statistical data

Aggregated statistical data will be obtained from the various data repositories (e.g. Eurostat dissemination data base); data will be arranged, analysed, stored and reported in an appropriate form.

2.3.b Secondary data in form of anonymised micro data sets:

Micro data sets will be obtained according to the applicable procedures. Rules set by the data providers (in particular, by Eurostat) for data storage, use, dissemination and the publication of results based on the analysis will be strictly followed.

2.3.c Primary data collection

Within the ENLIVEN project, primary data collection will predominantly involve the collection of qualitative data. Interviews will take place with the following categories of people (the categories are not mutually exclusive): policy makers; representatives of business interest organisations; representatives and members of trade unions and social movement organisations; participants in educational programmes; participants in organisational case studies (managers, line managers;

² <http://www.cedefop.europa.eu/it/events-and-projects/projects/european-skills-and-jobs-esj-survey/access-to-data>

³ http://www.europeansocialsurvey.org/about/structure_and_governance.html

HRD/HRM specialists; legal employee representatives; employees in their first 10 years of occupational careers). Particular emphasis will be given to interviews with employees in their early careers, as 3-5 young adults working for the organisations will be interviewed twice, with the goal of learning about their previous learning biographies and their current perception of learning opportunities available in the workplace. Records will take the form of interview recordings and transcripts.

Participatory observation and action research case studies are undertaken as part of WPs 5-7. WP 5 has an action research phase, in which small projects (learning projects, workshops, training) will be arranged with the members of the case study organisations. These will be tailored to the particular conditions of the organisation concerned, will run for up to 6 months and allow for on-site participatory observations for at least 4 working days. Records will be taken in the form of research notes made by the researchers involved in the field phase.

2.4 Data Use and Protection

In order to adhere to the principles of the Data Protection Directive and national legislation, and to the project's own ethical principles, the ENLIVEN consortium will apply the standards and procedures set out below. These are designed to minimise the risk of misuse of data by third parties, and to safeguard anonymity of research participants (in respect both of secondary use of existing datasets and data collected by the ENLIVEN project).

2.4.a Secondary analysis of EUROSTAT and related scientific-use micro data files

International quantitative micro datasets used (e.g. LFS, AES, PIAAC, ESS) are anonymised. None contain private or personal details regarding identifiable individuals. In the case of data made available by the European Statistical System, Eurostat imposes a strict framework to avoid misuse of (already anonymised) datasets. The key emphasis is on avoiding leakage of data to third parties who may break the anonymization by merging datasets with other datasets (for example, register data from the Social Security system).

In WPs 2-4, mainly⁴ scientific-use micro datasets provided by Eurostat will be used (LFS, AES, CVTS). When preparing the scientific-use files, Eurostat minimises the risk of disclosure of research participants' identities by removing information and by merging information into broader categories. It also requires research organisations applying to use these micro data to comply with detailed standards, which is, in this instance, the University of Leuven. Data will only be shared with ENLIVEN partners who were listed in the University of Leuven's data request. All partners will fully comply with the requirements⁵ set for the use of Eurostat's micro data files, in particular with regard to safe storage of micro data and safeguarding of respondents' identities in outputs based on the micro data used.

With regard to safe storage of micro data, ENLIVEN partners will:

- Ensure that they identify a senior researcher who will be responsible for storing the medium containing the confidential data in his/her office and will only allow access to them to the authorised researchers identified above.
- Ensure that research premises where the micro data are used and stored are secure.

⁴ Other data scientific use data sets might be also used (e.g. PIAAC, EJS, ESS) and rules set for their proper use fully observed.

⁵ Eurostat regulations/requirements - see in particular Section 8:
<http://ec.europa.eu/eurostat/web/microdata>

- Ensure that only authorised researchers who have signed a declaration of confidentiality have access to the micro data and that data are not copied in any way
- Store Eurostat Data-CD-ROMs and any information provided by Eurostat for data decryption in separate locked cabinets, making it unlikely that any third party can get both sets of information required for using the data sets.
- Take precautions to prevent any copies of the micro data being made, and to prevent the data being transmitted via the internet or within an organisation’s network (e.g. by using a stand-alone computer connected neither with the internet nor with the organisation’s local network). No printer will be attached to computers used for micro-data analysis.
- Document all use of the microdata set (who has used the dataset, on which dates, and for what purposes).
- Share data only as required amongst the members of the national research team undertaking a specific piece of research; that team will be responsible for keeping the data safe and for removing identifying information. Any such data will be shared only on this anonymised basis with the wider (international) ENLIVEN project research team.
- Ensure that all data, and other confidential items, are kept securely (i.e. accessible only to relevant members of the team), and that when such data and items are shared between partners, this is done in a secure way.
- Retain all data, and all other confidential items, on a secure password-protected site at the University of Nottingham (the “R: Drive”). Only members of the team will have access to this file store; it is managed by the Project Administrator (Ruth Elmer).
- Destroy datasets (and any output based on the datasets) in accordance with contractual requirements (e.g. for safeguarding non-disclosure of individuals, households, enterprises).

2.4.b Interview Transcription and Data Protection

The procedures applied for the processing of qualitative data in line with the requirements of data protection and ethical principles have been developed in detail in deliverable 11.2 and are summarised in Tables at 4.1 and 4.2 in the annex. They include:

- Safeguarding anonymity of research participants;
- Safeguarding confidential information;
- Restricting recording of sensitive information to the minimum required for answering the research questions and excluding any information from recordings which is either not relevant or potentially harmful;
- Ensuring fully informed consent from research participants.

Due to the difficulty in guaranteeing that recorded interviews will not contain personally identifiable data, such recordings will be subject to the Data Protection Act (or equivalent legislation in other countries). To ensure the security of the recording and its content, a research-institute-owned encrypted digital recorder or personally owned devices complying with the University of Nottingham’s policy and procedures relating to the use of mobile devices and remote working as set out in the Information Security Policy⁶ will be used.

Transcription will be carried out where possible within the premises of the research partners. In case of the use of external specialised service providers, the following rules will be applied. Recordings will be securely transferred to transcribers, for instance by secure upload to an approved transcriber’s

⁶ <https://www.nottingham.ac.uk/legalservices/documents/information-security-policy.pdf>

website. Encryption keys will be sent to the recipient by some other means, e.g. telephone, email, and will not be sent with the encrypted recording. Similar care will be taken with security of transfers of completed transcriptions from transcribers and to the safe return and/or secure deletion of recordings. Transcription agencies used will have a track record of academic transcription work and will be expected to sign undertakings confirming that the information is kept strictly confidential.

Informed consent forms will be distributed to all participants. They will be collected and stored in a locked facility, separate from the research material. Interviews will be analysed only in line with predefined research questions, as stated in the proposal and agreed on with the research participants in the informed consent process. The use of data for transcription purposes and the access to anonymised data sets for future projects will also be outlined in the informed consent forms.

Confidential or sensitive data will be kept in our password-protected data confidential ENLIVEN filestore, on the Research drive ("R: Drive") of the University of Nottingham's computer system. Only members of the team have access to this filestore. The University of Nottingham's Research Filestore is centrally managed and provided by University of Nottingham Information Services, and is subject to IS Business Continuity Planning, IS Information Security Policy and IS Change and Release standards. The service includes failover support across two separate data centres. Full backups to tape are performed once a week, while incremental backups are performed nightly. Tapes are retained for 112 days, after which they are re-used or securely disposed of. Access to the service requires authorised University of Nottingham credentials, and is restricted by the institutional firewall.

Data storage will comply with the UK Data Protection Act, equivalent national legislation, and applicable University of Nottingham and partner institutional policies. The Nottingham policy requires that data be retained intact for a period of at least seven years from the date of any publication which is based upon them. Data will be stored in their original form – i.e. tapes/discs etc. will not be deleted and reused, but kept securely.

2.5 Documentation

2.5.a Documentation Reports

Publications (e.g. books, articles in peer reviewed and professional journals, and presentations at scientific conferences), along with engagement with policy-makers and other public communities play a vital role in raising awareness about ENLIVEN results. To this end:

- All publications will be made available by open access.
- The project's web site, maintained by the University of Edinburgh, will have hierarchical levels for public engagement, interested stakeholders and detailed scientific information to increase and further participation and collaboration.
- Unless required otherwise by confidentiality undertakings given to research participants, data sets (anonymised as indicated below) used for analysis will be deposited on institutional data repositories and/or through the UK Data Archive at the University of Essex and/or through national data archives. (The Co-ordinator will approach the UK Data Archive with a view to making all datasets generated by the project available through that source at the end of the project.)

2.5.b Project Management Documents

All internal board, committee, work package and meeting/workshop reports will not be included in the ENLIVEN Dropbox folder, but will be stored on the University of Nottingham's system. Confidential items will be maintained in a password-protected confidential ENLIVEN filestore on the University of Nottingham's computer system. This will be accessible to members of all project teams, but will otherwise be confidential until and unless it is agreed that they should be made public.

On the basis of this management documentation, annual reports on the project's activities and progress will be prepared (using the quarterly Coordinators Reports used for Management Board Meetings). These reports will be approved by the Management Board. One report will be prepared which will be for public use (e.g. on website and for Advisory Board). However, if there is any sensitive or IPR-restricted information there will be a 'private' version too that this can be included in.

2.5.c Website

The web site is hierarchical in design with initial entry levels designed to promote the subject and the project in general, and H2020, in order to stimulate public engagement and to engage with interested stakeholders. In line with our data utility objectives, intermediate public levels will also be designed to support school children and their curricular activities. Higher open access research levels will detail scientific and technical progress as well as comprehensive information on dissemination activities.

In addition to providing links to each partner's respective normal research web pages, the web site will include various resources such as on-line publications, benchmark problems, discussion groups and links to other related sites.

2.6 Hardware and Software

2.6.a Intelligent Decision Support System

University of Nottingham resources and equipment (a desktop and a laptop) supported by funding from the ENLIVEN project will be utilised for the IDSS component of the project. IBM SPSS modeller software will be used. The University of Nottingham possesses a group licence for this software. Additional software usage will be covered by open source licences.

2.6.b Data Backup and Recovery

The University of Nottingham's Research Filestore ("R: Drive") is centrally managed and provided by University of Nottingham Information Services (IS), and is subject to IS Business Continuity Planning, IS Information Security Policy and IS Change and Release standards. The service includes failover support across two separate data centres. Full backups to tape are performed once a week, while incremental backups are performed nightly. Tapes are retained for 112 days, after which they are re-used or securely disposed of. Access to the service requires authorised University of Nottingham credentials, and is restricted by the institutional firewall.

2.7 Intellectual Property and Ownership

2.7.a Intellectual Property

The ENLIVEN partners will comply with intellectual property rights regulations and provisions in accordance with Horizon 2020 principles:

- All data generated will be centrally stored on the Co-ordinator's central computer system (University of Nottingham), using the University's standard archiving data control procedures and routinely (daily) backed up onto secure areas on a central server, in line with the above statement on data backup and recovery.

- The Project Administrator (Ruth Elmer) will hold management responsibility for these curation activities.
- All Consortium members are responsible for complying with the intellectual property regulations and provisions.
- At the earliest opportunity and appropriate times, data will be made publicly available by presentation at internal and external conferences and by publication in peer-reviewed journals. This is outlined in the ENLIVEN Dissemination plan.
- All data presented or published will be anonymised. Immediately after publication access to original anonymised data sets will be freely available on appropriate request, adhering to relevant regulatory requirements and ethical use of data approval. A condition of such access will be subsequent acknowledgment by those applying.

2.7.b Joint ownership

Joint ownership is governed by Grant Agreement Article 26.2 with the following additions:

Unless otherwise agreed: - Each of the joint owners shall be entitled to use their jointly owned Results for non-commercial research activities on a royalty-free basis, and without requiring the prior consent of the other joint owner(s), and each of the joint owners shall be entitled to otherwise Exploit the jointly owned Results and to grant non-exclusive licenses to third parties (without any right to sub-license), if the other joint owners are given: (a) At least 45 calendar days advance notice; and (b) Fair and Reasonable compensation

2.8 Open Access

Data with an acknowledged long-term value will be preserved and remain accessible and useable for future research. Data which supports and validates published research will also be preserved and, as far as possible, openly accessible to other interested researchers. All output, including academic publications, will be made available by 'Open Access' so that every paper will be available to readers without additional cost. Where Green route publishing is not possible, subject to resources authors will be provided with funding to publish through the Gold route.

All data presented or published will be in an anonymised form. Immediately post-publication access to original anonymised data sets will be made freely available on appropriate request, wherever possible through institutional data repositories. Release of such datasets will be in accordance with relevant regulatory requirements, especially on the ethical use of data. A condition of access and reuse of the data will be subsequent acknowledgment of ENLIVEN as a source.

2.9 Quality Assurance

ENLIVEN has an annual Quality Assurance System reporting plan which details the project deliverables each partner is responsible for delivering, and the peer reviewers responsible for checking them. The document also details the timescales for each deliverable, including the formal date for submission and internal timelines. The latter include dates for first drafts to be sent to the Co-ordinator by the Responsible Partner, the deadline for the Nottingham team to distribute to all Partners, the deadline for feedback from all Partners and the date for feedback from Peer Reviewers. Certain protocols and procedures are also outlined: e.g., when documents are sent to all Partners, this is for *comment only*, mainly in respect of accuracy in relation to statements about partners' countries or specific research input. Feedback must be given in the body of an email (*not* via word 'tracking' on the document itself as this creates a very high volume of work). Partners have three working days to respond with any comment; otherwise, correctness is assumed. Finally, the *peer reviewer offers detailed*

review/amendment regarding overall coherence and quality (with reference to the Quality Criteria document) - and comments and changes can be made in word tracking.

3 Work-Package-specific Data Issues

This section sets requirements for data management related to specific work-package tasks.

3.1 WP1: Mapping European and national policies and programmes, and their contribution to economic and social inclusion

This WP maps and investigates, at European, national, and subnational levels, **policies and funding schemes** to tackle disadvantage, inequality and social exclusion. It includes a **comparative study** of policies and programmes in selected European countries and Australia, and analyses the role system characteristics (policy regimes) play in participation and inequalities. The main methods used are critical discourse analysis of **policy documents and funding schemes**. Country-based analysis of selected education programmes involving **in-depth case studies** (based in part on interviews) will collect data about individuals' employment (e.g., ex-ante and ex-post, working conditions), empowerment (e.g., self-perceptions of personal and collective agency) and active citizenship (e.g., political participation, participation in community life). Learners will be selected for interview according to programme, funding received, orientation, and provider (public/private).

3.2 WP2: Constraints and facilitators of access and participation

This WP **analyses** national institutional architecture and qualifications frameworks **using policy documents and key informant interviews**. Within each participating country, we will produce a **map of the institutional architecture** with a view to assessing its flexibility and inclusivity, based on policy documents and reviews, including Eurydice reports. All partners will analyse aggregated administrative and survey data, national and local policy documents, and qualifications frameworks, to illuminate institutional architecture in their jurisdiction. Up to ten **key informant interviews** will be conducted in each jurisdiction **with policy-makers, service providers and service users** to provide **commentary** on the extent to which official accounts are consistent with the practical experience of service users and those involved in delivery of services. **Interviews with young adults** (some of whom will be potentially vulnerable adults) will explore barriers to participation among various groups.

3.3 WP3: The role of European governance in adult education & learning policy

This WP **maps the main actors** contributing to adult education policy developments in Europe. It identifies key policy actors at global, regional, national and sub-national levels via a thorough **web search**. It improves understanding of the coordinating function of European governance in adult education by identification and examination of governance mechanisms coordinated by or under the supervision of EU shared institutions, how they work, and how they have developed. It also looks at evidence of these governance mechanisms, and related EU policies, influencing public and regulatory agencies intervention in adult education markets at national and sub-national levels. It also **analyses the development of taxonomies and indicators in European adult education** and their **use of cross-national survey data (e.g. PIAAC)** to inform policy. Finally, it examines how taxonomies and indicators have developed in European adult education, and interrogates how PIAAC connects with governance mechanisms under the supervision of EU shared institutions.

3.4 WP4: Improving our understanding of the effect of system characteristics by building stronger data and adding a longitudinal, regional & sectoral focus

This WP constructs a **pseudo-panel data set on lifelong learning participation**, starting from **the time series of the Labour Force Survey (LFS)**, in order to investigate how lifelong learning participation and system characteristics have developed over time. LFS also permits breaking-up of the sample by NUTS-region (Nomenclature of Territorial Units for Statistics) and by employment sector. In order to improve develop **richer contextual information**, the LFS-based pseudo-panel data set will be enriched with **more detailed information from other datasets** (e.g., AES and PIAAC). To facilitate longitudinal, regional and sectoral analysis, indicator time series will be constructed to be linked to the panel data set. New indicators will be developed for characteristics that have not yet been fully covered, such as indicators of the demand for general, sector-specific and job-specific skills. However, micro data sets will not be merged in a way to avoid any risk of potentially compromising the anonymity of any individual units represented.

3.5 WPs5–7: Studying the role of workplace learning and patterns of work organisations for early career structuration; qualitative interviews on learning biographies

WP 5 and 6 utilise **in-depth organisational case studies**. 16 in-depth organisational case studies will be undertaken, drawing on 64 interviews with managers, line managers, HRD experts and legal representatives of employees, and 64 in-depth interviews with adults in their first 10 years of their occupational career will be conducted. Each organisational case study will include two interviews with each of four managers, HRM/HRD professionals and line managers and at least one interview with an employee interest representative. During an **action research phase** small ‘learning projects’ will be arranged with the members of the case study organisation, tailored to the particular organisation, and will run for up to 6 months. In WP6, some potentially **sensitive data** will be collected (for example, **personal reports on observed conflicts in the workplace**); the young adults (aged 18-35) will be invited to share important aspects of their learning biographies, focusing on job-related non-formal and informal learning in the workplace.

In WP7, beyond a broad review of the relevant literature, **Case studies** will be implemented in three countries (ES, AT, SK) on three initiatives advocating better employment conditions during early career stages, including young employees among their activists. Case studies will be based on two expert interviews (one with an official speaker for the initiatives, one with a representative of a corresponding employer interest organisation) and one interview with a young activist. Where feasible and not running against the outlined ethical principles, the activist interviewed will be employed by the organisations studied in WP5 and 6. For the interviews with the young activists, particular safeguards will be taken to shield them from any potential harm resulting from their participation in the research. Further detailed data management in these workpackages is provided in Table at 4.3 below.

3.6 WP8: Knowledge discovery on evidence-based policy making in participating countries; & WP9: Establishment of Intelligent Decision Support System for evidence-based policy making

This work package involves knowledge discovery based on data available from international organisations (e.g., European Commission, OECD, UNESCO) and other ENLIVEN research findings to establish a **knowledge based methodology** (Case Based Reasoning: CBR) designed to facilitate analytical insights, informed governing of ‘policy problems’, and modelling of policy making on ALE/LL

programmes in the EU. A knowledge based system requires knowledge acquisition, which will begin with an in-depth analysis, working with other WPs, of **existing data** (e.g. ESS, LFS/SILC, Eurydice, Cedefop and Eurostat) and case studies. Based on the analysis, data mining techniques will identify case representation; investigations will then be conducted based on case representation and training in the case base to identify similarity measure models. These different stages are conducted iteratively, until the performance of the CBR system provides satisfactory evaluation metrics (accuracy and user feedback, etc.). Establishing an IDSS is an iterative process. Findings from WPs1-7 will contribute to **building common models** of policy making, reflected in the IDSS procedure.

3.7 WP10-11: Dissemination and Project Management & Integration

A **data repository**, comprising anonymised data gathered during the project, and main findings, will be maintained for further use. This will include transcripts of interviews as well as data collected for case studies. On publication of a related output, or on conclusion of the project, these data will be transferred from the University of Nottingham's Research Drive to the University's Research Data Management Repository (<https://rdmc.nottingham.ac.uk/>) and/or to the UK Data Archive (<http://www.data-archive.ac.uk/home>).

Where feasible, partners will also deposit their data in institutional or national data repositories. This will be done in strict accordance with European and individual country Data Protection legislation, and in the light of the standards and guidance developed by the Consortium of European Social Science Data Archives (CESSDA: <https://cessda.net/>). This will include ensuring that permission has been given by individuals and organisations that contribute data in any form not already public. All stored data will be thoroughly anonymised (except in the case of expert interviewees who explicitly decide not to be anonymous).

The data will be citable and will be given a Digital Object Identifier code (DOI) within the data repository which will be guaranteed for 7 years, in line with Table 2, below.

4 Appendices

4.1 Table 1: Key elements of the framework to ensure anonymization within the ENLIVEN research process (for storage/use within the project): (to be refined within the research project)

Level	Item	Description	When
1	Removing names/using replacements	Family name (removed); forename (replaced by a name typical for the gender, generation and the community and not the name of another person in the sample surveyed); replacing the name of the employing organisation; replacing the name of the geographic entity (town, etc.)	Immediately, when transcriptions are made.
2	Removing indirect identifier	Transforming information on educational attainment, age, migrant background of a particular kind, household characteristics into categories (e.g. 25 to a 25-29 age brackets).	When including transcripts in the R: Drive.
3	Removing information allowing an indirect identification	Removing information which allows the identification of a person by her/his particularities, habits etc. for an observer familiar with the particular context.	When choosing information for publication (biographical vignettes, quotations); before preparing interview transcripts for archiving (if any).
4	Alienation and strategic replacement of important information	Replacing important (“telling”) details which potentially include a risk of de-anonymization by related details capable of communicating comparable information.	When presenting information in biographical vignettes.

4.2 Table 2: Processing of data in the qualitative research implemented by the ENLIVEN project

	Expert Interviews; Interviews with representatives of business interest organisations and trade unions/social movement organisations	Qualitative Interviews with participants in educational programmes & with participant in organisational case studies	Data collection within the participatory observation/action research phase of the case studies
Informed consent	Forms on informed consent will be collected and stored on the R: Drive separate from other research material.	Forms on informed consent will be collected and stored on the R: Drive separate from other research material.	Informed consent forms for participants in the action research phase will be collected and stored on the R: Drive separate from other research material
Collection	Face-to-face interviews or phone interviews; on-site interviews in surveyed organisations	Face-to-Face interviews (all); on-site case study enterprises (managers, line-managers); neutral environment at choice of the interviewee (all others)	Participatory observation/participation in the action research module
Recording	Personal identifiers (name, age, function, related information): paper and pencil; experts are free to decide whether their names/the names of their organisations should be reported or should be anonymised; main interview: digital voice recording (a numeric code links the voice recording to the personal identifier)	Personal identifiers (name, age, function, related information): paper and pencil; main interview: digital voice recording (a numeric code links the voice recording to the personal identifier)	Research Diary for on-site participant observation – paper and pencil; all personal identifiers will immediately be replaced by codes representing particular research participants (or omitted, where not relevant for the research activity)
Organisation and storage	When participants choose anonymity or are regarded as vulnerable: personal identifiers (single paper copy); information will be stored in a locked compartment. Information on personal identifiers will be collected and stored electronically only in an anonymised version, attributing pseudonyms, using broader categories (e.g. age); two digital copies of the main interview will be stored in the R: drive separate from other research material; copies will be stored in a locked compartment separate from the papers with the personal identifiers; interviews will be transcribed (verbatim transcription); passages where personal	When participants choose anonymity or are regarded as vulnerable: personal identifiers (single paper copy); information must be stored in a locked compartment. Information on personal identifiers will be collected and stored electronically only in an anonymised version, attributing pseudonyms and using broader categories (e.g. age); two digital copies of the main interview will be stored in the R: drive separate from other research material; copies will be stored in a locked compartment, separate from the papers with personal identifiers; interviews will be transcribed (verbatim); passages where personal identifiers of the interviewee or related	Notes must be stored in a locked room and a locked compartment

identifiers of the interviewee or related persons (e.g. name of a co-worker) are revealed will not be transcribed, but will be immediately replaced in a way safeguarding anonymity [e.g. “the interviewee”; “an interviewee’s co-worker”]; individual transcripts will be further anonymised by the responsible researcher prior to compiling the transcripts within the software package used. Transcripts will be held only within the originating partner research organisation and the University of Nottingham R: Drive.

persons (e.g. name of a co-worker) are revealed will not be transcribed, but will be immediately replaced in a way safeguarding anonymity [e.g. “the interviewee”; “an interviewee’s co-worker”]; individual transcripts will be further anonymised by the responsible researcher prior to compiling the interview transcripts within the software package used; only anonymised interviews will be saved in the R Drive. Transcripts will be held only within the originating partner research organisation and the University of Nottingham R: Drive.

<p>Adaption and alteration</p>	<p>When participants choose anonymity or are regarded as vulnerable: personal identifiers will be anonymised; verbatim transcripts will be anonymised in a two-step procedure (immediate replacement of identifiers during transcription; further removal of any information helping to identify the interviewee prior to compilation of interviews; a summary of each interview will provided [10,000 signs] and translated into English; for the summary, any further information which risks disclosure of respondent identity will be removed/replaced by more general information; at this stage, any sensitive information (e.g. on health condition, on sexual orientation etc.) not required for the research process will also either be removed or replaced by a more neutral statement (e.g. a description of a concrete case of misconduct will be replaced by a statement that a case of misconduct has been reported).</p>	<p>Personal identifiers will be anonymised; verbatim transcripts will be anonymised in a two-step procedure (immediate replacement of identifiers during transcription; further removal of any information helping to identify the interviewee prior to the compilation of interviews; a summary of each interview will provided [10,000 signs] and translated into English; for the summary, any further information which risks disclosure of the respondent’s identity will be removed/replaced by more general information; at this stage, any sensitive information (e.g. on health condition, on sexual orientation etc.) not required for the research process will also either be removed or replaced by a more neutral statement (e.g. a description of a concrete case of misconduct will be replaced by a statement that a case of misconduct has been reported).</p>	<p>Researchers will produce a detailed summary of their notes as an electronic document for each day of observation, but removing all personal identifiers; potentially harmful information/sensitive information not relevant for the project must not be included in the summaries</p>
<p>Retrieval and consultation</p>	<p>Interview transcripts (with alterations required for anonymisation) will be stored [in original language] within the software used for qualitative text analysis, separately in each partner’s organisation (and saved on the University of</p>	<p>Interview transcripts (with alterations required for anonymisation) will be stored [in original language] within the software used for qualitative text analysis, separately in each partner’s organisation (and saved on the University of Nottingham R: Drive);</p>	<p>(Anonymised) summaries of on-site visits/participatory observation will be shared within the local research team</p>

	Nottingham R: Drive); otherwise, within the partnership, only translated interview summaries will be exchanged and compiled in one file (this will be the base for cross-country comparative analysis).	otherwise, within the partnership, only translated interview summaries will be exchanged and compiled in one file (this will be the base for cross-country comparative analysis).	
Use	Interviews will be analysed only in line with predefined research questions, as stated in the proposal and agreed with the research participants in the informed consent process.	Interviews will be analysed only in line with predefined research questions, as stated in the proposal and agreed on with the research participants in the informed consent process.	Interviews will be analysed only in line with predefined research questions, as stated in the proposal and agreed on with the research participants in the informed consent process.
Dissemination	Original language interview transcripts will not be disseminated but will remain with the team of the research organisation responsible for the research. Fully anonymised transcripts will be saved on the University of Nottingham R: Drive. They will also be made available along with summaries (in English, and from which all information potentially disclosing the participants' identities is removed) at the end of the project on a suitable data archive.	Original language interview transcripts will not be disseminated, but will remain with the team of the research organisation responsible for the research. Fully anonymised transcripts will be saved on the University of Nottingham R: Drive. They will also be made available, along with summaries (in English, and from which all information potentially disclosing the participants' identities is removed) at the end of the project on a suitable data archive.	Research notes will not be disseminated, but will remain with the team of the research organisation responsible.
Alignment and Combination	Expert interviews will be analysed in combination, however, no further risks to confidentiality of information are expected	Interview transcripts of respondents from each organisation will be analysed in combination; special attention will be paid to the risk of disclosure of respondents' identities when bringing the information from the various interviews and information from the participatory observations together; information which carries the risk of disclosure will be omitted	Summaries of research notes will be combined for each organisation studied and merged with the transcripts of the interviews; special attention will be paid to the risk of disclosing respondents' identities when material is brought together; information which carries the risk of disclosure will be omitted
Deletion/ Destruction	Data will be retained intact for a period of at least seven years from the date of any publication which is based upon them. Data will be stored in their original form – i.e. tapes/discs etc. will not be deleted and reused, but kept securely. Except	Data will be retained intact for a period of at least seven years from the date of any publication which is based upon them. Data will be stored in their original form – i.e. tapes/discs etc. will not be deleted and reused, but kept securely. Except where	Personal field notes (paper and pencil) will be destroyed at the end of the project (on acceptance of the final report).

where individually agreed by the research participant, all records which could lead to identification of individual persons will be destroyed at the end of the project (on acceptance of the final report by the Commission).

individually agreed by the research participant, all records which could lead to identification of individual persons will be destroyed at the end of the project (on acceptance of the final report by the Commission).

