Work Package 3

Data Mapping Framework

Deliverable 3.1

MapIE - Framework for Data mapping

Delivery date: 28.2.2025 (latest update 2.10.2025)

Dissemination level: Public

Lead Beneficiary: Tampere University (TAU), University of Helsinki (UH)

Authors: Marjanen Jukka, Koskinen Antti, Lintuvuori Meri, Nyman Laura, Asikainen Mikko & Vainikainen Mari-Pauliina

Funded by the European Union

Disclaimer

The content of this deliverable reflects the authors' views. The European Commission is not liable for any use that may be made of the information contained in the present document.

Copyright message

Table of Contents

Version history

Version	Date	Updated by
1.0, first draft	13.6.2024	J. Marjanen, M. Lintuvuori, A. Koskinen, M. Asikainen, M.-P. Vainikainen
		Comments and discussions with all partners
2.0, second draft	14.11.2024	J. Marjanen, A. Koskinen, M. Lintuvuori, L. Nyman, M. Asikainen, M.-P. Vainikainen
		Comments and discussions with all partners
3.0, final draft	24.1.2025	J. Marjanen, A. Koskinen, M. Lintuvuori, L. Nyman, M. Asikainen, M.-P. Vainikainen
		Draft submitted for internal review
3.1, for submission	21.2.2025	J. Marjanen, A. Koskinen, M. Lintuvuori, L. Nyman, M. Asikainen, M.-P. Vainikainen
3.1.1, revised after D4.2.	25.6.2025 17.9.2025 2.10.2025	J. Marjanen, M. Lintuvuori, L. Nyman, M. Asikainen, M.-P. Vainikainen

Glossary of acronyms

Acronym	Extended definition
DoA	the Description of the Action
MapIE	Mapping of Longitudinal Data of Inequalities in Education
IRT	Item Response Modelling
SES	Socio-economic status
SEN	Special educational needs
ISCED	International Standard Classification of Education

1 Basic information about the project

MapIE is a 4-year project funded by the European Union through the Horizon Europe framework programme. The project has a total budget of 2, 699,644.75 € divided among 7 beneficiaries. The associated partner Szegedi Tudomanyegyetem from Hungary receives funding from their national funding authority. Tampere University (Finland) acts as a coordinator for the project. Table 1 summarises general information about MapIE.

Table 1. Key facts about MapIE project

Title	**Mapping of Longitudinal Data of Inequalities in Education**
Acronym	MapIE
Grant Agreement No.	101132474
Funding Programme	HORIZON-CL2-2023-TRANSFORMATIONS-01
Instrument	HORIZON-RIA
No. of Beneficiaries	7
No. of Associated Partners	1
Project start date	1.3.2024
Project end date	29.2.2028
Project duration	48m

2 Rationale for developing inclusion criteria for datasets

In the MapIE data mapping framework the objective is to identify datasets that make it possible to study how educational inequalities develop and identify effective mechanisms for closing achievement and well-being gaps. The metadata of the identified individual datasets will be combined into a well-defined and curated pool of multiple metadata files, referred to as the MapIE Data Catalogue. The rationale for undertaking this effort is to:

1) Facilitate the discovery of datasets that would otherwise require in-depth searches or personal contact with the authors of the datasets.
2) Facilitate answering of research questions about the development of educational inequalities and formulating scientific arguments about mechanisms for overcoming these educational inequalities, neither of which can be answered with individual datasets alone.

3 Inclusion criteria for datasets

3.1 A brief overview of criteria

In order to create a well-defined and curated MapIE Data Catalogue, a set of criteria has been developed. These criteria aim to ensure that the datasets described in the Catalogue enable a better understanding of how inequalities develop and of effective mechanisms for closing achievement and well-being gaps. In the criteria, requirements are set for 1) how learning outcomes are assessed longitudinally, 2) how the sample is designed, 3) which factors have been used for educational inequality, 4) where the data have been collected and 5) whether the principles for ethical science have been adhered to.

If the dataset meets the inclusion criteria, the metadata of the dataset can be included in the MapIE Data Catalogue, even if the dataset contains additional measures/data. The inclusion criteria and their rationale are discussed in detail below. The citations refer to extracts from the Description of the Action (DoA)[1] of the MapIE project and call texts presented after the list of criteria.

[1] *The Description of the Action (DoA)* is the Annex of the Grant Agreement which contains the details of how the project will be carried out.

3.2 Detailed criteria with rationales

Criteria 1: Longitudinal learning outcome measurement

The metadata of the dataset can be included in the MapIE Data Catalogue only if the dataset contains:

- At least two measurements of the same pupils' learning outcomes of academic performance, or knowledge development (e.g., mathematics, literacy, science, problem-solving, critical thinking).
- The measurements of learning outcomes for individual time-points must use either 1) the same test for all measurement points and ages, 2) equated tests (using anchor items or other scaling methods), or 3) age-group normative tests applied consistently across all pupils (these may be different for different age groups/grades, as long as the test for a single measurement point is the same for all pupils of the same age).
- At least two measurements of learning outcomes are from ISCED levels 1-2 or from the transition stage to or from these levels. If transitions are covered, one of the measurements must still be from ISCED levels 1 or 2.
- The minimum interval between the first and the last measurements of learning outcomes is at least seven months within a school year.
- The most recent measurement of the learning outcomes must be from 2010 or later.

The metadata of the dataset will NOT be included in the MapIE Data Catalogue if the only measures of learning outcomes are teacher-given school grades based on tests that are not comparable. However, teacher-given grades included in the datasets can be utilized in

The rationale:

The project scope, as defined in the call and the DoA, focuses on "learning outcomes," "learning progress," "education achievement" and "educational inequalities." (Call1; Call2; Call3; DoA1). Therefore, datasets must include measurements of at least one dimension of pupils' learning outcomes (e.g., academic performance, knowledge development) while potentially including other measurements such as motivational factors as covariates.

The call and DoA emphasize the importance of tracking individual pupils' progress and learning trajectories, necessitating a pupil-level longitudinal design (Call6; Call7; Call8; Call9; DoA5, DoA6). This design is essential for understanding how educational inequalities develop and evolve at the individual level over time. Consequently, the dataset must include measurements of the same pupils' learning outcomes over time.

At least two comparable measurements of learning outcomes are crucial as they enable distinguishing between temporary fluctuations and lasting changes in educational outcomes. This longitudinal approach allows for analysing how different factors influence learning trajectories and how educational inequalities develop or diminish over time (Call9; DoA7; DoA8). The requirement for at least two comparable measurements represents the minimum needed to establish a baseline and measure change. However, the dataset can include several measurements of the learning outcomes, as well as other measurements.

Key learning outcomes need to be measured using 1) the same test, 2) equated or 3) age-group standardised tests to allow comparison within time points. Scale equation allows measures from different data collection instruments (e.g. tests) to be treated as equivalent even if their content is not exactly the same. This can be achieved, for example, by using anchor items and item response modelling (IRT). This ensures that observed changes represent genuine learning progress rather than measurement artifacts. Age-group standardised tests can for instance be tests used for specific age-groups in national assessments even if they were corrected by teachers. The tests for different grade levels or years do not necessarily need to be linked as long as the same test has been used for the entire cohort or large-scale sample. Within each measurement point, all individuals in the age group receive the same test instrument, but the instrument may differ between measurement points. The instrument must remain unchanged throughout a single measurement point. This ensures that observed changes represent genuine learning progress rather than measurement artifacts.

While the core learning outcomes must meet these comparability requirements, datasets may include additional measurements using other approaches. Datasets in which the only measure of pupils' learning progress are school grades given by teachers are excluded, if learning outcomes are measured using different tests and grades are based on information that varies between schools and teachers.

As educational inequalities develop gradually, a minimum of 7 months is required between measurements within the same school year. This allows the inclusion of datasets with measurements from the beginning to the end of a school year, while excluding those that mainly cover the summer holidays. Datasets may include additional measurements with shorter intervals as long as at least one pair meets the 7-month criterion. On the other hand, the interval between measurements may be much longer than 7 months. This approach maintains high standards of data quality while maximising the inclusion of valuable longitudinal datasets.

According to the call and the DoA (Call4; Call5; Prop2; DoA3; DoA4), the focus is on ISCED 1 (primary education) and ISCED 2 (lower secondary education) and at least two measurement points are required from these levels or the transition stage to or from them (e.g., school readiness measured at the end of kindergarten or test for entering ISCED level 3). If transitions are covered, one of the measurements must still be from levels 1 or 2. Additional measurements may come from other ISCED levels. This focus ensures coverage of crucial stages of development when many educational inequalities emerge and consolidate.

The requirement for the most recent measurement to be from 2010 or later ensures relevance to contemporary European educational contexts, capturing recent influences of educational systems, policies, technological advances and socio-economic conditions (Call5). As previous measurements can be prior to 2010, this criterion allows for the examination of long-term inequality trajectories and the analysis of changes in contributing factors over time.

Criteria 2: Sampling design criteria

The metadata of the dataset can be included in the MapIE Data Catalogue only if **either**:

1) The dataset is based on a random sample of the target population. The random sample may be, for example, a simple random sample of pupils, a cluster sample of schools, or some other type of probability sample. The sampling design can use stratification.
2) The dataset aims to cover the entire target population (missing pupils/schools are allowed).

The target population can be a municipality, a larger administrative area (e.g. a state or a geographical area) or the whole country.

The metadata of the dataset will NOT be included in the MapIE Data Catalogue if the dataset only uses:

- Convenience samples, where participants are selected based on their availability and willingness to participate.
- Voluntary samples, where participants are self-selected.

The rationale:

The sample design criteria are set to ensure that the datasets included in the Catalogue reflect the true characteristics, diversity, and distribution of the populations being studied. With either a random sample or a full population in the data, statistical techniques and inferential methods can be confidently applied, and results can be generalised to the wider population. This makes the datasets more suitable for informing educational policy and making educational decisions. The sampling criteria are also designed to ensure that very small datasets of, for example, two or three schools participating in an intervention are not included in the Catalogue.

Criteria 3: Criteria for inequality factors in the data

The dataset can only be included in the MapIE Data Catalogue if it contains variables/information that can be used to study at least one of the following dimensions of potential educational inequality:

- SES inequality (e.g. family income or parents' education level)
- Ethnic inequality (e.g. immigrant background or ethnic minorities within countries)
- Gender inequality (e.g. gender / sex)
- Geographical inequality (e.g. urban/rural or segregation of residential areas within cities)
- Special educational needs (SEN) or support needs (e.g. tier-level support not based on diagnoses or support needs in the language of teaching)
- Inequalities related to the education / school-system (e.g. tracking, school selection, public/private schools)

The rationale:

The above criteria have been linked to educational inequalities in previous academic literature. As the focus of the longitudinal measurements is on individual pupils, the datasets should include at least some pupil-level variable that is potentially related to educational inequalities. In addition, they were identified as key factors in educational inequalities in the WP2 literature review. They are also mentioned in the call and in the DoA (Call4; DoA9; DoA10; DoA11). Consequently, the included datasets must contain information on at least one of these dimensions.

Criteria 4: Location criteria

The metadata of a dataset can be included in the MapIE Data Catalogue only if:

- The dataset is by an organisation located in Finland, Norway, Sweden, Germany, Luxembourg, or Hungary.

The rationale:

The geographical scope of the MapIE Data Catalogue is deliberately focused on datasets from either Finland, Norway, Sweden, Germany, Luxembourg, or Hungary. Firstly, understanding the educational context is crucial for analysing the mechanisms of inequality. Only the participating countries can provide comprehensive contextual descriptions that allow us to identify the underlying mechanisms of educational inequalities.

Secondly, these specific countries were chosen strategically to represent different educational approaches within European systems. The inclusion of the Nordic countries (Finland, Norway, and Sweden) provides an insight into the "Nordic model" of education, which is characterised by its later tracking. In contrast, the Central European countries (Germany, Luxembourg, and Hungary) represent systems with earlier tracking practices. This intentional diversity of educational systems allows us to examine how different structural approaches to education might influence the development of educational inequalities.

Restricting the Catalogue to these countries ensures that all the datasets included can be analysed in the light of thoroughly documented educational contexts and policies. This approach allows for more nuanced and accurate interpretations of how educational inequalities develop within specific systemic contexts, and how different educational approaches may contribute to or mitigate these inequalities.

Criteria 5: Ethical criteria

The metadata of the dataset can be included in the MapIE Data Catalogue only if the authors of the dataset certify that ethical rules and laws were followed in the collection of the data.

The rationale:

The project originally responsible for collecting the dataset has followed the ethical rules and legislation that were in force at the time of data collection (DoA12). This information is collected as part of the data mapping survey, where the owner of the dataset declares that ethical and legal rules were followed during the data collection process.

3.3 References to the EU call and the Description of the Action (DoA) of the MapIE project

Call refers to extract from the EU call. DoA refers to the extract from the Description of the Action of the MapIE project.

Call1: "…analysis of the available literature and a selection of techniques used to assess inequalities in education, training and learning achievements over time."

Call2: "Identify the interventions that compensate inequalities in learning outcomes over time…"

Call3: "It is even more difficult to examine any causal link between educational policies and inequalities in education achievements…"

Call4: "Proposals should map and collect surveys with a longitudinal design, following the learning progress of students over time, and linking with information for example on students' achievement scores, socio-economic background, language spoken at home, migration history, the school they attend, or whether they have attended early care and childhood education."

Call5: "Proposals should enhance the knowledge base of what shapes the educational outcomes over time, how to best support the learning of all students and reduce education inequalities in the short, medium and long-term."

Call6: "Map and collect existing regional or national longitudinal data allowing to follow individual cohorts of students over time…"

Call7: "…longitudinal studies of individual student cohorts are rare, and it is difficult to follow students, assess their learning outcomes over time…"

Call8: "Proposals should map and collect surveys with a longitudinal design, following the learning progress of students over time…"

Call9: "…follow students, assess their learning outcomes over time and identify trends in education inequalities."

Call10: "…we look at system-, regional- and school-level policies and practices and individual-level factors (e.g. socioeconomic status, immigrant background, gender, special educational needs) that are associated with the development of educational inequalities…"

DoA1: "…identify means for compensating inequalities in learning outcomes…"

DoA2: "The prevention of gendered selection of further education and the labour market begins at the basic education level and these perspectives will also be considered in the empirical analyses conducted in this project."

DoA3: Table 1 datasets: 21 of 27 described datasets have at least two measurement points from ISCED 1 to ISCED 2 levels.

DoA4: "Our project targets especially basic education, which according to the ISCED standard comprises primary education (ISCED1, first stage of basic education) and lower secondary education (ISCED2, second stage). Almost all the population of learners in basic education are enrolled in recognised education (table 3). Besides basic education, we also look at the transitions from pre-primary education and early care to basic education, and from basic education to ISCED3 level upper secondary education…"

DoA5: "As stated in the scope of the call, we map all available studies following the learning progress of pupils over time…"

DoA6: "…little research is published based on pupil-level longitudinal empirical data that give comparable information about how inequalities develop…"

DoA7: "We aim at understanding how educational inequalities develop as an interplay of pupil-level predictors and the features of the local, regional and national policies and practices."

DoA8: "With already available data, we are able to describe the mechanisms of the development of educational inequalities…"

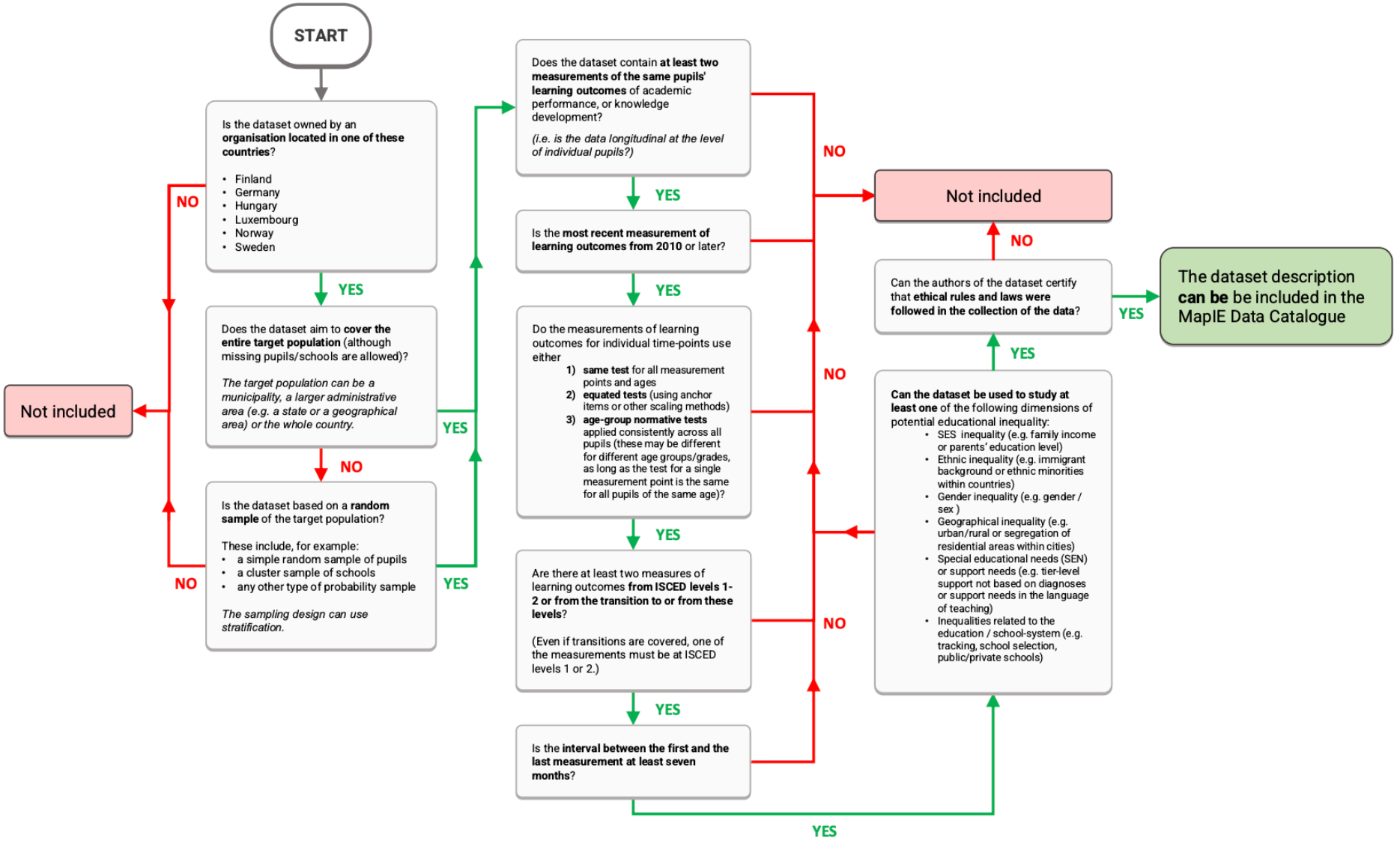
DoA9: "Our project focuses particularly on understanding the contextual differences between the Nordic and Central European countries, as historically they have had different approaches to tracking and school selection…"

DoA10: "…equal possibilities for learning for every child regardless of their socio-economic or ethnic background or the area they live…"

DoA11: "School segregation and tracking are usually quite strongly related with pupils' background."

DoA12: "We will check whether the original projects that have been responsible for collecting and pseudonymising the data have undergone official ethical reviews and in the empirical analyses, only use data that meet all the principles of ethically conducted research. We also include information about research ethics in the public descriptions of the data in the metadata database."

3.4 Flowchart for dataset inclusion criteria



4 Identifying sources of local variation

The structure of the education system and the specific characteristics and societal challenges of the local context influence both the definitions of equality/inequality and the phenomena and factors that are considered as relevant when measuring them or implementing interventions to compensate for them. Therefore, in WP3 we also describe the contextual factors that cause and influence educational inequalities and the policies to address them.

For this purpose, country descriptions including the administration, structure and organization of the education system are done for the target countries in the MapIE project (Finland, Norway, Sweden, Germany, Luxembourg and Hungary). The focus is on which structural and organizational factors in their current state and in recent reforms have seen to create equality as well as inequalities, particularly in terms of students' background characteristics, geographical location, gender, socio-economic status (SES), ethnic background and special educational needs (SEN).

The country-specific descriptions cover especially the ISCED levels 1 and 2 which comprises basic education in primary and lower secondary level, which are compulsory for all. However, the descriptions also consider factors related to transitions within and between ISCED levels that may contribute to inequalities in terms of the transitions from pre-primary education and early care to basic education, and from basic education to ISCED 3 level upper secondary education.

Each partner systematically collects information from multiple sources in their own countries to describe all relevant contextual considerations. One potential challenge related to the methodology of this WP is to find sources to describe the systems and their features in a comparable way since the main aim of the study is not to conduct comparative policy research based on document analysis. To overcome this challenge, we also utilise existing relevant descriptions, such as EU resources describing education systems and student demographics across EU countries (i.e. Eurydice, Eurostat and the European Agency for Special Needs and Inclusive Education). In addition, locally selected references and studies are used to highlight phenomena and factors that are important to study in that particular context.

For example, studying educational equality from the perspective of immigrant children is a relevant question everywhere, but there are big differences in the number of immigrants in different countries or parts of them. Different countries also have different minority groups to consider. For example, the education of Roma children is an important equality issue to be studied in Hungary, whereas in the Nordic countries, the research on

this particular subgroup is quite marginal. There may also be policies and interventions addressing inequality that are not clearly identified as effective practices in previous research, but are nevertheless implemented on the basis of theory, political intuition or preference (e.g. bussing, initiatives to engage parents of children with low SES, provision of for local homework support, language adaptations, free school meals, etc.). Thus, data collected in different local contexts may have different indicators depending on the context. These contextual factors must be taken into consideration not only in the mapping and analysis of data, but especially in the dissemination of results when policy makers discuss such issues.

The country descriptions are collected in the template documents (Excel spreadsheets, tables, descriptions) in the MapIE WP3 and can be updated and extended as needed. These will serve as background data both in data mapping conducted in WP4 and data analyses of WP5. A systematic analysis of the country descriptions will be written in a scientific book chapter within WP5.

References

European Commission. (n.d.). *Eurydice*. https://eurydice.eacea.ec.europa.eu/

European Agency for Special Needs and Inclusive Education. (n.d.). *European Agency for Special Needs and Inclusive Education*. https://www.european-agency.org/