

Scottish and German Perspectives on Access to Longitudinal Research Data in Education

Christian Aßmann, Michael Bergrab, Nathan Burns,
Tunde Csoban, Saikat Ghosh, Louise Kelly, Ryan Kirkpatrick,
Ryan Stewart, Friederike Schlücker, Doris Stingl

Juli 2026

Scottish and German Perspectives on Access to Longitudinal Research Data in Education

Christian Aßmann^{1,2}, Michael Bergrab^{1,*}, Nathan Burns³, Tunde Csoban³,
Saikat Ghosh¹, Louise Kelly³, Ryan Kirkpatrick³, Ryan Stewart³,
Friederike Schlücker¹, Doris Stingl¹

July 2026

<https://doi.org/10.5281/zenodo.21369783>

¹ Leibniz Institute for Educational Trajectories (LIfBi), Germany

² Chair of Survey Statistics and Data Analysis, Otto-Friedrich-Universität, Germany

³ University of Strathclyde, Glasgow, Scotland, United Kingdom

* Corresponding author

Abstract

The research potential of available data hinges severely on the underlying survey mode and data collection design, as well as the institutional settings to access available data for empirical research and survey planning. We provide a review of different data resources available to researchers in the domain of educational research and of the institutional settings for data access in Scotland and Germany. Criteria to gauge the relative advantages of different designs and institutional settings for data access relate to longitudinal population coverage, linkability, time readiness, findability, accessibility, interoperability, and reusability. The relative advantages are contextualized in (arche-)typical situations for survey planning and possible improvements are discussed.

Keywords:

Research data access, Research data infrastructures, Longitudinal educational data

1. Introduction

The availability of high-quality research data is essential for analysing educational processes and outcomes, evaluating policy measures, and understanding social inequalities. In this context, research data refer to data made accessible for scientific analysis, typically to members of accredited universities or non-university research institutions, i.e. the scientific community. Both Scotland and Germany maintain well-established infrastructures for educational data collection and scientific data access, though they differ markedly in governance, scope, and degree of centralization. This paper situates Scotland and Germany within a broader global landscape of educational panel and assessment infrastructures. Numerous empirical studies in education and the social sciences rely on longitudinal or repeated cross-sectional survey infrastructures that enable the analysis of learning trajectories, social mobility, and competence development over time. At the international level, prominent examples in the education domain include the Programme for International Student Assessment (PISA), which assesses the competencies of fifteen-year-olds across OECD member countries every three to four years (OECD, 2023), and the Programme for the International Assessment of Adult Competencies (PIAAC), which measures adult literacy, numeracy, and problem-solving skills in cycles of roughly ten years. At the national level, longitudinal designs have become integral components of educational research systems. In Germany, the National Educational Panel Study (NEPS) traces learning biographies from early childhood to adulthood (Blossfeld & Roßbach, 2019), in Scotland, the Growing Up in Scotland (GUS) study follows children and their families across key life transitions (Growing Up in Scotland, 2025), and in the United Kingdom, the Avon Longitudinal Study of Parents and Children (ALSPAC) provides rich, multi-generational data on health, education, and social outcomes (University of Bristol, 2025). Data from these studies have been used to investigate the educational and competence trajectories, the transitions within the education system, and the sustained educational inequality, see e.g. Fuchs and Wößmann (2007), Niemann et al. (2017), Sixt and Aßmann (2020), Matthewes (2020), Passaretta and Skopek (2021), Gilbert et al. (2021), Volodina et al. (2023), Smyth and Privalko (2023), Ghosh and Kleine (2024), Sawyer et al. (2025), Aßmann et al. (2024), and Burns et al. (2025) to name only a few.

A key dimension of research data relates to the underlying participation regulations during data collection. Administrative data typically arises from (law-based) mandatory participation governed by legal obligations, for instance, when data are collected for operational purposes by public authorities.¹ While these binding design regulations ensure broad longitudinal coverage of the population, their use for research purposes is typically secondary and depends on the availability of access infrastructures, data documentation, and appropriate regulatory frameworks to unlock their full research potential.

¹ An example is Art. 79 Sozialgesetzbuch IV in the context of German social insurance or the National Dental Inspection programme which mandates every child should receive an oral health inspection in their first of year primary school.

In contrast, survey data are usually based on voluntary participation based on informed consent.² However, mixed participation regulations are also common, especially in large-scale educational assessments such as PISA or the German Bildungstrend studies, see OECD (2023) and Stanat et al. (2023), where participation in cognitive tests may be mandatory for schools and pupils, but additional contextual data are collected via voluntary background questionnaires. Participation and consent regulations are closely related to participation rates, where informed opt-in consent is in line with non-mandatory participation and opt-out consent in line with mandatory participation typically reflected in higher participation rates and better population coverage for the latter. In general, opt-in regularities potentially threaten the longitudinal population coverage, see among others Zinn et al. (2020).

These participation frameworks and informed consent practices are also applicable to the context of data linkage. In Germany, survey respondents may consent to the linkage of their survey data with administrative records using identifiers such as name, birthdate, or a consistent pseudonym if available. Consent rates in longitudinal panel surveys are typically high (70% – 90%), reflecting the trust-based relationships built over repeated data collection waves (Antoni & Schmucker, 2019; Bachbauer et al., 2021). In Scotland, consent to link data can come from the individual or via approvals from the organisational unit that own the data. Administrative records in Scotland typically have high population-level coverage (>90% of population). Beyond individual-level linkage, geographic linkage plays an important role, too. In Germany, this often takes place at the level of municipalities, whereas in Scotland geographical areas (e.g., data zones, council areas, health boards) serve as the standard linkage unit.³ Statistical analysis generally benefits from linkage at the individual level or for smaller geographical units, allowing better coverage of individual and geographical heterogeneity and relevant regional contexts. Law based linkage of administrative data suited for research purposes is also possible, where in Germany the legal framework for this is currently in a state of upheaval with new potentials for linking being possibly mandated by the envisaged law on improved data use for research currently under discussion, see also Bundesministerium für Forschung, Technologie und Raumfahrt (2025) and German Data Forum (RatSWD, 2025, 2026).

Access to research data is in general governed by institutional infrastructures designed to ensure compliance with legal regulations most often relating to data protection issues. In Germany, this involves a decentralised system of research data centres (RDCs), which regulate access according to the Federal Data Protection Act (BDSG) and the European General Data Protection Regulation (GDPR).⁴ The degree of anonymization of data and the mode of access,

² Corresponding regulation is provided via the European General Data Protection Regulation (DSGVO art. 6 Abs. 1, 1, a and f) or the UK General Data Protection Regulation and Data Protection Act 2018.

³ Data zones in Scotland are geographical areas designed to cover similar-sized proportions of the population. Data zones can vary in size, but each contains around 800 people. In contrast, municipalities in Germany are administrative units varying significantly in terms of covered population and area. Furthermore, characteristics at the level of postal codes are typically not available in Germany

⁴ As an example in the German context, access to administrative data on employment biographies for research purposes was established in 2004 alongside labour market reforms implemented between 2003 and 2005 to allow for valid monitoring and research-based evaluation of policy measures, see Müller and Möller (2019a).

e.g. download, remote, or on-site use, depend on the risk of reidentification as assessed by the data provider. In the United Kingdom (UK), access is regulated by the Data Protection Act 2018 and the UK GDPR, which largely mirrors its European counterpart but includes specific exemptions for research in the public interest. Scotland's centralised education system and the presence of institutions such as the Research Data Scotland support integrated data infrastructures and promote harmonised access to administrative and survey data.

The comparison of Germany and Scotland offers a unique opportunity to examine how different national contexts shape the governance, collection, and use of research data. Both countries are embedded in broader European data governance frameworks while following distinct traditions in education policy, federalism, and institutional arrangements. Scotland is a world leader in data linkage with multiple different data controllers such as Public Health Scotland (PHS) or electronic Data Research and Innovation Service (eDRIS). As part of the UK, Scotland benefits from additional data infrastructure of the Office for National Statistics (ONS) and Administrative Data Research UK (ADR UK). Moreover, Scotland has also developed a research-ready data environment through initiatives such as the Research Data Scotland. These efforts include secure access models such as the Safe Haven framework, accompanied by strong ethical and legal oversight, see Gao et al. (2022) and Research Data Scotland (2025c) for an overview. In contrast, Germany's research data ecosystem is characterised by a decentralised and federal structure. Access to administrative and survey data is facilitated by a network of legally independent research data centres with informal coordination strengthened by the German National Research Data Infrastructure (NFDI) established since 2020. While this model supports data sovereignty at the state as well as institutional levels and emphasises legal compliance with data protection laws, it can result in fragmented access conditions, varying documentation standards, and limited possibilities for data linkage across domains, see Connelly et al. (2016) and RatSWD (2025) for the limitations of administrative data and access thereto. Understanding these structural and cultural differences is crucial for the development of comparative perspectives on data access, data linkage, and ethical governance in educational and social research.

This article compares the institutional and legal frameworks governing access to mainly educational research data in Germany and Scotland. While the focus is on longitudinal data, we also consider relevant cross-sectional datasets that enable data enrichment through linkage. The comparison is organised along five analytical dimensions: (a) survey design and sample access, (b) sample sizes and representativeness, (c) thematic scope and measurement, (d) consent and linkage possibilities, and (e) data access, metadata and reusability. These dimensions describe the empirical and institutional characteristics of the data sources under review. In addition, we relate the respective data infrastructures to the FAIR principles – (F)indability, (A)ccessibility, (I)nteroperability, and (R)eusability – as formulated by Wilkinson et al. (2016). In this context, FAIR compliance is understood not as a property of the substantive content or analytical purpose of a dataset, but as a feature of the infrastructures, documentation standards, metadata practices and access arrangements that enable data and metadata to be found, accessed, linked, interpreted and reused over time. By distinguishing

between dataset characteristics and infrastructure-level conditions for FARIness, the article contributes to debates on ethical, legal and technical aspects of research data governance in the social sciences.

The remainder of this article is structured as follows. Section 2 gives an overview of the institutional contexts regarding school systems and the data ecosystem in Scotland and Germany, section 3 describes key features of Scottish research data sources, section 4 outlines the German data infrastructure, section 5 compares both systems and discusses potential improvements, and section 6 concludes.

2. Institutional contexts: school systems and data ecosystems in Scotland and Germany

In the following, information on the institutional contexts are provided in terms of the school systems as the place where institutionalized education happens and the data ecosystem enabling access to research data.

2.1. The school system in Scotland

Since 1999, Scotland has been given limited self-governing powers to make laws on devolved matters under the Scotland Act 1998 (Scottish Parliament 2024), including education. Scotland is divided into 32 local authorities, with the Education (Scotland) Act 1980 outlining the provision of education for each authority (Scottish Government 2019). Young people in Scotland complete 7 years at Primary School (P1-P7), beginning at the age of 4 or 5, followed by up to 6 years of Secondary school (S1-S6). Secondary school is compulsory in Scotland until the age of 16 (S4) (GOV.UK). The education curriculum in Scotland, Curriculum for Excellence, splits education into two phases: the Broad General Education (BGE) and the Senior Phase (Education Scotland 2025). The BGE focuses on young people from pre-primary school until their third year of secondary school (S3), with the final three years of secondary school (S4-S6) collectively forming the Senior Phase (Education Scotland 2023).

The professional judgement of teachers is the key to the assessment of students throughout the BGE (Scottish Government 2024). Low stake online assessments facilitated by The Scottish National Standardised Assessments and the Measaidhean Coitcheann Nàiseanta airson Foghlam tron Ghàidhlig are undertaken as part of typical classroom activities in P1, P4, P7, and S3. These assessments are to provide an insight on the numeracy and literacy abilities of the young people and understand their strengths and development needs. The evaluation of young people's progress is never made under the assessment outcomes, with the key measure of progress under BGE being the data collection of teachers' judgement on the annual Achievement of Curriculum for Excellence levels.

Young people are assessed in a range of subjects through three levels of official exams during the Senior Phase: National 5, Higher, and Advanced Higher (McEnaney 2025). Typically, a young person will first sit National 5 exams in their fourth year of secondary school (S4), with this level

of exam able to be taken at any point during the Senior Phase. Successful young people at National 5 level can progress onto Higher level, usually first sat in their fifth year (S5). Young people can sit additional Highers in their sixth and final year (S6), or progress onto Advanced Higher if successful at Higher. Progression to the next level is decided at a school-level, with young people often needing a B or a C at the previous level (Dow 2026).

Post-secondary education includes, but is not limited to, university and college (often referred to as Higher Education and Further Education, respectively) (mygov.scot 2026). Entry to university typically requires at least four Highers from the one set of exams, with the most competitive programmes such as medicine and dentistry often requiring more Highers as well as Advanced Highers (Circuit 2024). Many university courses which do not require an Advanced Higher often still encourage young people to obtain one, as the well-respected qualification can prepare young people for university-level learning (Sgoil Lionacleit 2025). College courses are often shorter than university courses, offering many post-study options (Scotland). Young people who study at college can often be allowed direct entry to a related university course, granting entry to the second or third year of the programme. This is referred to as articulation (University of Strathclyde). Young people studying their first degree at either university or college are eligible for free tuition, with the Student Awards Agency Scotland paying five years of tuition fees (House of Commons Library 2024). This allows a young person to complete up to a four-year honours degree through a typical or articulation path, with an additional year for a repeated year of study or a course change if necessary.

2.2. The school system in Germany

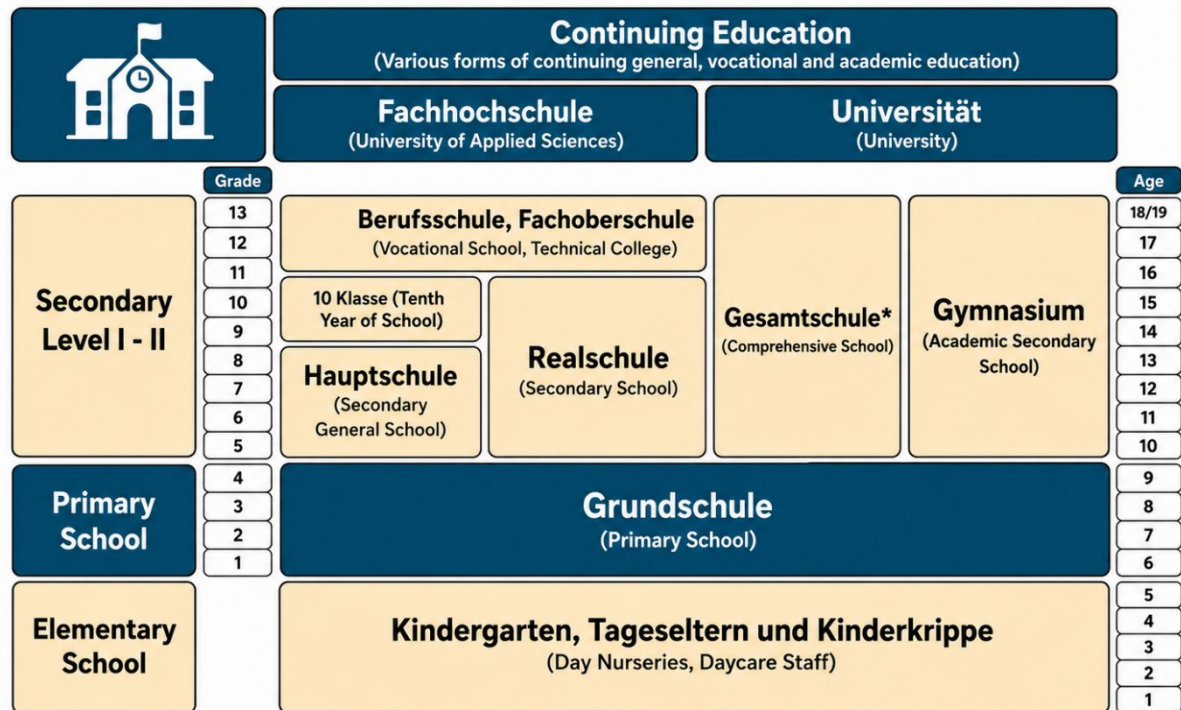
Germany's school system is shaped by federalism. Educational legislation and administration are primarily the responsibility of the federal states, the *Länder*, particularly in the school sector. Coordination across the *Länder* is pursued through the Kultusministerkonferenz (Standing Conference of the Ministers of Education and Cultural Affairs). This arrangement creates a common national framework, but substantial institutional, curricular and administrative variation remains across the *Länder*.

Compulsory schooling generally begins in the year in which children reach the age of six and usually comprises nine years of full-time schooling, with some *Länder* requiring ten years. Primary education is provided in the *Grundschule* where pupils are usually aged six to ten.⁵ A defining feature of the German system is the differentiation of lower secondary education into educational pathways and school types. The main pathways include the *Hauptschule*, *Realschule* and *Gymnasium*, although these may be organised in different institutional forms across the *Länder*, including single-track schools, schools combining several tracks and comprehensive schools. Lower secondary education can lead to different certificates such as the *Hauptschulabschluss* and the *Mittlerer Schulabschluss*, which are commonly associated with subsequent vocational education and training. Access to the *gymnasiale Oberstufe*, by

⁵ Although in Berlin and Brandenburg the primary school extends to age twelve (six years instead of four years).

contrast, provides the route towards the Abitur and thus to higher education. For an overview of the German school system see figure 1.

Germany's Education System



Visuals by
Studying-in-Germany.org



Figure 1: Germany's education system. Adapted from "Germany's Education System" by Studying-in-Germany.org, licensed under CC BY-SA 4.0. Changes were made to improve readability and accessibility. The adapted figure is distributed under CC BY-SA 4.0.

Germany therefore represents a more institutionally differentiated school system. Although formal permeability between school types and tracks exists, educational trajectories are structured earlier and more explicitly around school-type or track distinctions. This institutional differentiation is central for comparative research because school type, transition points and federal state level regulation can shape both educational trajectories and the construction of comparable samples.⁶ For the present article, the federal structure is particularly important because it affects both research design and data access. Sampling strategies must often account for all Länder, school types and transition regimes, while legal and administrative requirements may vary across the Länder.

2.3. The data ecosystem in Scotland and Germany

The data ecosystems in Scotland and Germany differ in ways that closely reflect their institutional settings. Educational research does not only depend on the availability of

⁶ For more details see among others <https://www.kmk.org/en/bildungsministerkonferenz/bildungsthemen/educational-paths-and-qualifications.html>.

individual datasets, but also on the institutional ecosystem through which data are produced, linked, documented and accessed. In this article, the term data ecosystem refers to the interaction between administrative data production and collection, statistical reporting, research data centres, linkage infrastructure and access governance. This distinction is important because the same substantive research question may be much easier to address in one country than in another, because the institutional conditions for access and linkage differ.

In Scotland, the education-data ecosystem is closely connected to a comparatively integrated administrative structure. Local authorities and schools provide pupil-level information to national data collections, and the Scottish Exchange of Data (ScotXed) supports secure electronic data exchanges between the Scottish Government and wider service communities. Core administrative sources include the Scottish Pupil Census and related school statistics on pupil characteristics, attendance and absence, exclusions, attainment, qualifications and leaver destinations. These data provide a strong basis for population-level educational monitoring and, where governance conditions allow, for longitudinal research using linked administrative records.

A central feature of the Scottish ecosystem is its linkage orientation. Research Data Scotland and ADR Scotland provide metadata and access routes for curated administrative datasets, while eDRIS and the National Safe Haven support secure access to sensitive or linkable data for approved research. Access is not open in the sense of unrestricted data download; rather, it is organised through controlled procedures, public-benefit assessment, researcher accreditation, disclosure control and secure research environments. This makes the Scottish case particularly relevant for research questions that require population coverage, stable administrative identifiers and the possibility of linking educational information to other domains such as health, social care or later transitions. At the same time, the Scottish ecosystem remains shaped by sector-specific coverage and governance boundaries. Administrative data are strong for institutional trajectories and outcomes, but less flexible for measuring subjective attitudes, family processes or behavioural mechanisms, and individual-level coverage is not equally integrated across all educational sectors.

Germany's data ecosystem is organized differently. School education up to the end of secondary education is primarily the responsibility of the Länder, and administration in the school sector is almost exclusively the responsibility of the Länder. This federal structure shapes not only school organization, but also data production, harmonisation and access. Official school statistics are available at national and federal state levels, including information on pupils, school types, vocational schools, foreign language instruction, special educational support, repeaters and school leavers. However, these official statistics mainly support monitoring and reporting. They do not amount to a single, integrated national pupil-level research infrastructure comparable to a centrally linkable administrative register.

At the same time, Germany has a highly developed research-data infrastructure for educational research. Key resources include the National Educational Panel Study (NEPS) or German samples from international school performance such as PISA, TIMSS, the IQB Trends in Student

Achievement, and administrative labour-market data provided by the Institute for Employment Research. These resources are prepared, documented and made available for scientific use through specialised Research Data Centres. They are complemented by research-data networks and by selected linkage products, most notably NEPS-ADIAB, which links NEPS survey and competence data with administrative labour-market records from the IAB on the basis of informed consent.

A more precise comparison is therefore not that Scotland possesses educational data whereas Germany lacks them, but that the two countries differ in the institutional organisation of data access, linkage and reuse. Scotland can be characterised as having a comparatively integrated and linkage-oriented administrative data ecosystem, whereas Germany's education-data infrastructure is more federally distributed and institutionally specialised. The German case is marked by considerable strengths, including rich longitudinal survey data, competence measures, large-scale assessment data and well-established research data centres. At the same time, the federal organisation of schooling and the division of responsibilities across federal states, domains and institutions create more dispersed arrangements for data governance, legal authorisation and linkage. In this sense, the German data ecosystem is best understood as fragmented not because of a lack of high-quality educational data, but because access to and linkage between such data are mediated by a more complex and decentralised institutional architecture.

3. Research data and access in Scotland

The Scottish Government has committed itself to improving Scotland's future via high-quality data, analysis, and evidence to inform policy and support and improve service delivery. Scotland has developed strategies to better utilise its existing data while fostering a culture, where legal, ethical, and secure data linkage is accepted and expected. As part of this, the Scottish Government has multiple databases available for access from many different sectors, including communities and the third sector, economy, education, health and social care, and money and tax (Scottish Government, 2025f). These data vary from population-level to individual-level. There are 115 national individual-level datasets available in Scotland, which are accessible to researchers. These datasets are mainly related to health outcomes, e.g., hospital admissions or access to health services, however, also include data related to education. The Pupil Census, in detail, collects data from all publicly funded schools, covering topics such as pupil enrolment, class sizes, attainment, and other data related to teachers, early learning and childcare, see Section 2.1. The Longitudinal Education Outcomes data for Scottish Universities holds information on earnings, employment, and benefits linked to the education records of students attending Scottish universities (Scottish Government, 2025b). However, these data are currently not available for linkage. Table 1 and table 2 summarise and compare the educational data infrastructures and the findings of longitudinal data sources in Scotland and Germany along with the FAIR criteria.

Prior to 2024, there were multiple different ways to access national data in Scotland. This process was often unclear, complicated by the necessity to engage with multiple different organisations, each with their own process. Although this was time-consuming, data access and linkage was still possible. In 2024, the Research Access Service (RAS) was launched by Research Data Scotland in partnership with the Scottish Government, with the aim of making data access faster and simpler where there is a recognised benefit to the public good (Research Data Scotland, 2025d). The service was developed in collaboration with recognised stakeholders, i.e. researchers, analysts, and policymakers, and is currently in its beta phase. The RAS currently controls access to nine of the most popular Scottish national datasets, with access to the remaining 106 datasets being provided through the Public Benefit and Privacy Panel (PBPP) pathway, organised by Public Health Scotland (PHS). During the access process, see figure 2, the researcher is allocated a research coordinator, who guides them through the process, from first inquiry to end of the research study for which data was required. The RAS framework provides a single entry and end point for the two pathways. During the initial enquiry, the research coordinator advises the researcher whether they are expected to follow the RAS or PBPP pathway. While these datasets are not free to access, cost estimates are provided upon inquiry, before any committal to progressing with the data access project, but each dataset typically costs around £6,000. Within Scotland there are seven main data controllers, each linked to either the Scottish Government or an university (Scottish Government, 2025f). Currently, approved organisations that can request access to data are limited to UK-based public sector institutions, such as selected universities, the National Health Service (NHS), local authorities, and the Scottish Government. Researchers from other organisations, including commercial and industry, can partner with an approved institution or directly with the data controllers, who could conduct analyses on their behalf (Research Data Scotland, 2025d).

Access pathways to data in the National Safe Haven

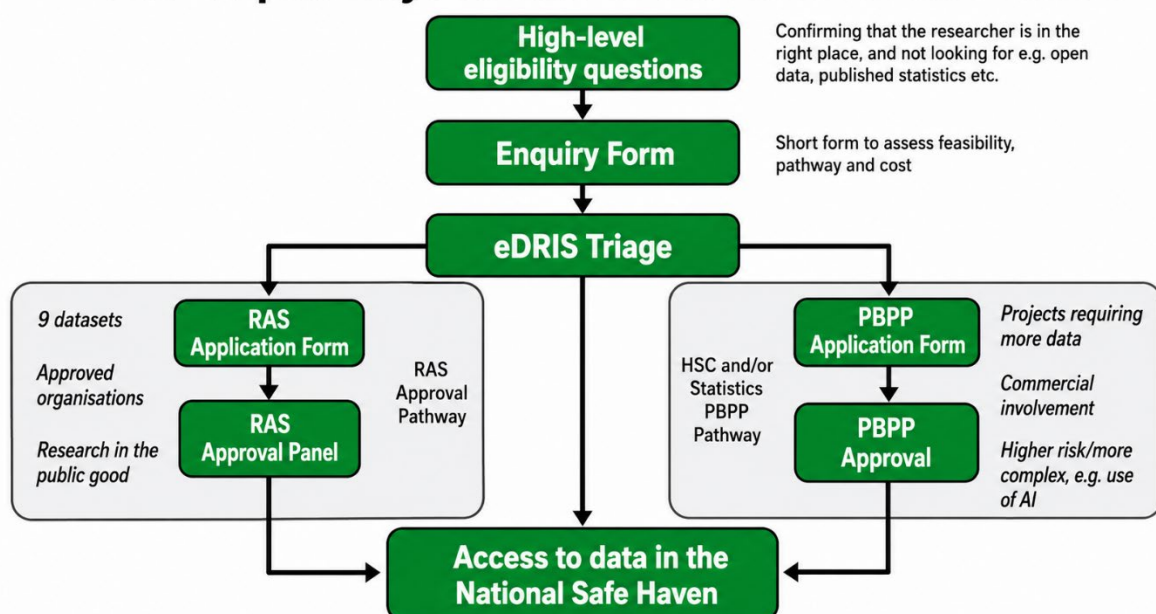


Figure 2. Access pathways to data in the National Safe Haven. Own illustration based on Research Data Scotland's applicant guidance for the Researcher Access Service, eDRIS guidance on PBPP applications and data linkage, and Scottish Government documentation on Safe Havens and governed access to health and social care data (Research Data Scotland, 2025a; Research Data Scotland, n.d.; Scottish Government, 2023).

As well as access to national datasets, researchers may also benefit from linkage of the different datasets to expand the number of variables and hence the research questions covered by the data. Data linkage does not always require approval directly from the individual but can be granted via the RAS or PBPP framework described previously. All research considered by RAS or PBPP must have a benefit to the public good to receive ethical approval. Data linkage has long been used in Scotland with protocols and guidance in place. Linkage is undertaken by the electronic Data Research and Innovation Service (eDRIS) who manage the data on the behalf of the data owners. eDRIS support the access to data from inquiry to end of the research project. Typically, given the sensitive nature of the data, access is provided via Trusted Research Environments (TREs), sometimes known as secure data environments. TREs provide a secure area for researchers to access the data and complete the analysis using a range of licensed software. These secure environments and measures ensure that the public can be confident their data remain secure while contributing to research in the public interest. Currently, there are four physical TREs in Scotland (Glasgow, Edinburgh, Dundee, and Aberdeen), each of these are linked to both a university and an NHS health board. In these TREs, the researcher is provided with a room equipped with a computer to access the data and these can be booked by researchers who have approval to access the data. Researchers may also request to access the data remotely via the National Safe Haven. In this case, data can be accessed on a secure laptop, within a secure environment (e.g., private office), by using a virtual private network (VPN). The latter approach has the benefit of not having to travel for data access and analysis can be done in the researchers own time. Analysis must be conducted using licensed software, and raw data cannot be taken out of the virtual safe haven environment. All output which is to be shared must go through disclosure checks to ensure there is no identifiability in the data. This can include no cell count less than 5, or no geographical count less than 10.

Data linkage of health records has been possible in Scotland for many years due to the Community Health Index (CHI) number. The CHI number is a unique identifier for all individuals who engage with the National Health Service in Scotland. The identifier, which is ten digits long, is made up of an individual's date of birth (6 digits), two random digits, a single digit for the sex of the individual (even if female, odd if male), and a final arithmetic check digit. A person's CHI number follows them through life and provides the ability to identify a person's health records across all sectors, including general practice, general dental practice, and hospital care. In addition, each pupil who attends school in Scotland has their own identifier (pupil ID) and those who take assessments in Scottish schools receive a unique identifier called the Scottish Candidate Number. In cases where there is no unique identifier available at time of contact, personal identifiers (typically forename, surname, sex, date of birth, home postcode) could be linked via probabilistic methods. The process involves indexing, which restricts the

comparison of records to plausible pairs, such as restricting to individuals from the same council area only, individuals with the same name, or individuals of the same sex. Each identifier is compared to each plausible match, which is typically taken from a complete list of all individuals. Each comparison produces a score on the agreement or disagreement weights based on how common the rate of agreement is where matches on rare features (e.g., names) score more and matches on common features (e.g., sex) scoreless. A composite score (the sum of all weights) is then calculated and compared to a given threshold to determine if the record is a match. This threshold is continually reviewed using a set of performance metrics e.g., precision, recall, and clerical review. PHS aim for precision greater than 99% at the expense of some recall to minimise false matches.

3.1. School-Pupil Census

The School-Pupil Census is an annual Scottish Government survey, issued to obtain accurate data on publicly funded schools, for example regarding local authority, grant-aided, school centres, and their pupils. Pupils in Scotland must attend school by law from 5 to 16 years old, with pupils typically attending primary education from age 5 to 11, and secondary education from age 12 to 16. After 16, pupils may choose to do a further one or two years at secondary school, or may leave to attend college, begin an apprenticeship, or join the labour market. Some pupils in Scotland will attend special need schools. These schools are typically designed for children who require additional support through a recognised disability.

The census collects information on two levels, i.e. school and individual. Each school is allocated a seed code, a unique identifier for the school which can be used to link across different school-related datasets. At the school-level, data are collected on school type categorized as primary, secondary, and special, class sizes and number of teachers in terms of full-time equivalents. At the pupil-level, the data are forename and surname, home address, date of birth, sex, ethnic background, student stage ranging from first primary to sixth secondary and special, free school meal registered, additional support needs, and absences (bi-annually). Information on the survey used for data collection, and a full list of variables is available online, see Scottish Government (2025e). Data access is available through the PBPP pathway as described above. Data are first collected for an individual pupil at the time they are registered at primary school. Each year, the school will send a request to parents or carers for updates. Each year, the Scottish Government designate a specific day to be 'Census Day'. It is by this date that all schools must provide relative information back to the Scottish Government, via ScotXed (an online database), through the data collection system (procXed). During data upload, each dataset is validated and if any issues are identified, these are immediately flagged. Once the data have been validated, they will then be transferred to the centralised database.

The Scottish Government publish annual summaries of the School-Pupil Census survey aggregated at the school level. These are released as two datasets: School Contact Details (Scottish Government, 2025c) and School Summary Statistics (Scottish Government, 2025d). The former is updated multiple times per year at irregular intervals, while the latter is published annually, usually in March of the current academic year. For example, the School Summary

Statistics 2024 release, which contains data on students beginning their registration in the 2024/2025 academic year, was published in March 2025. These datasets are not managed by the data controllers described above, they are publicly available to the general, international public and can be used for research purposes, at no cost, under the UK Government's Open Government License (UK Government, 2025). As of 5th May 2025, School Summary Statistics starting from academic session 2011/12 can be accessed and downloaded via the Scottish Government websites (Scottish Government, 2025c). Prior data could be obtained via a Freedom of Information (FOI) request to the Scottish Government.

Both datasets contain the total population of state-funded schools in Scotland, but not of independent schools, as they are not required to disclose information to the School-Pupil Census. The School Summary Statistics 2024 release contained 2419 unique schools with 1978 primary, 360 secondary, and 107 special schools, see Scottish Government (2025d), which did not include the 91 independent schools also open at the time, see Scottish Government (2025a). A complete list of the variables included in the School Contact Details and the School Summary Statistics is shown in Tables 1 and 2, respectively. Examples include each school's address, postal code, and local authority and which schools have been opened and/or closed since January 2006. Variables can be added, removed, or updated between releases, although the list of included variables has been mostly standardised since 2019.

The School Contact Details and School Summary Statistics data can be linked to one another and to other datasets, using the school's unique identifier variable (Seed Code) or via the school's name and postal code. The advantages of these aggregated data over the individual-level School-Pupil Census data are that they are easy to access at no cost, the data are complete, except for independent schools, which make up only 3.8% of the total schools, and are frequently released and updated. The latter factor in particular means that researchers can utilize these data, for example in planning surveys, knowing that the lag-time is less than one year. This would allow for poststratification of the survey responses validated on the most recent release of the summary statistics. Further school-level data may be requested via FOI requests. This includes, but is not limited to, school attainment data obtained via FOI to the Scottish Qualifications Authority. All information released via FOI is then published, see Scottish Qualification Authority (2025).

Data at the individual level can be accessed via the process described in section 2. This can be used to link records to other national datasets or surveys. To be successful in an application for data access, researchers must show sufficient evidence that the outcome of the research is to the benefit of the public, where this is determined via clear aims and outcomes. Further, access to individual-level data requires ethical approval. This is most done through the RAS application, see section 2. However, in some cases, it may be appropriate for researchers to seek further ethical approval from their institution. Universities may have their own ethics processes, procedures and boards; however, it is also possible for faculties or departments to set up their own within the university structure to provide this extra approval. Individual level data which can be accessed in this way include the previously discussed health data linked to

CHI numbers, pupil data linked to pupil IDs and also birth registrations and further data collected in educational settings. These data can also be linked using probabilistic methods at an individual level. Aggregated results can then be shared for publications at differing levels with strict protocols, e.g. no cell can contain a count of less than 5, or geographically less than 10.

3.2. Longitudinal Education Outcomes

Longitudinal Education Outcomes data for Scottish Universities was launched in September 2025. This dataset contains information to explore research questions related to university education and careers covering 2005/06 to 2021/22 (Office for National Statistics, 2025). Currently, it is nonlinkable, however, the data is derived from the Higher Education Statistics Agency and is linked to information on employment, earnings, and benefits. Further information is available on the industrial classification of employment. The owners of this data (Office for National Statistics) plan to add more data as and when available, although there is no information available on whether this will be linkable in the future. If it was to be linkable, there would be the option to follow a child from birth, through school, and into the job market for those who attended a Scottish university. Currently, questions that could be answered relate, for example, to labour market outcomes controlling for the attained university degree.

3.3. Birth registrations

Data are available via the process described in section 2 on all children registered in Scotland following birth. This dataset can be used to form cohorts and identify children for sampling purposes. The data are also used to publish annual population estimates by administrative area (e.g., council area, area-based deprivation) and age. National Records Scotland also provide weekly births registrations for Scotland overall. These are published with a one-week delay, for example, births registered in week 1 of the year, are published at the end of week 2. Monthly birth registrations are also published at council area and health board level. These are published with a one-month delay, for example, monthly births for September are published by the end of October. All these publications include open access data of records at differing aggregation levels (typically council area and health board) with yearly statistics further split by sex and age.

3.4. Data Collection in Educational Settings

Scotland has utilised the educational setting to collect data related to health outcomes. This includes oral health and body weight and height data which are used to monitor trends in these health conditions and refer to services where appropriate. National inspections are conducted in every local authority school in Scotland and, in the case of oral health, are statutory right for children (UK Government, 1978, 1980). These typically work on opt-out consent basis. Data are collected on a few topics and can be accessed via the process described in section 2. These data could be linked to the school pupil census or other educational data e.g., attainment.

3.5. Growing up in Scotland

Growing Up in Scotland (GUS) is a nationally representative longitudinal cohort study that has followed Scottish children from early childhood into adolescence and beyond since 2005, see Growing Up in Scotland (2025). Drawing its samples from the Child Benefit Register, the study achieves near-complete population coverage at baseline and employs regular face-to-face interviews with the child's main carer, supplemented by self-completion instruments as children grow older. Access to the data is provided through the UK Data Service, while sensitive variables such as geographical identifiers or linked administrative data require secure-access arrangements. Initial cohort sizes range from 3,000 to 6,000 children, with attrition remaining moderate across sweeps and still allowing for robust subgroup analyses, for example by socio-economic status, region, or family structure. The thematic scope is broad and policy-relevant, covering children's cognitive and socio-emotional development, health, childcare and schooling, parenting practices, financial circumstances, neighbourhood context, and social inequalities. Standardized instruments – including Strengths and Difficulties Questionnaire (SDQ)⁷ and various cognitive assessments – ensure comparability across waves and with other UK studies. GUS operates a clear, staged consent procedure for linkage to administrative records, including health, education, and local authority data, thereby ensuring transparency and adherence to Scottish governance requirements. Data documentation, metadata and user guides are comprehensive, and most datasets are readily accessible for research, with sensitive data provided in safe settings. Overall, GUS performs strongly regarding the FAIR principles: it is easily discoverable, well documented, methodologically transparent, interoperable through harmonized measures, and offers a high degree of reusability due to its longstanding, stable study infrastructure.

3.6. Linkable dataset from the Programme for International Student Assessment

A recent development in Scotland is the inclusion of a linkable dataset from the Programme for International Student Assessment (PISA) within the ADR Scotland Programme. Through this initiative, accredited researchers can now request access to Scottish-specific PISA datasets for 2012, 2015, and 2018. These data are held in the Scottish National Safe Haven and can be linked to a wide range of administrative education datasets, including the Pupil Census, Attendance, and Exclusions records (Research Data Scotland, 2025b). This linkage potential significantly strengthens opportunities for longitudinal and cross-dataset analyses within the Scottish context. This centralised infrastructure therefore reduces administrative burden and enhances the potential for largescale, policy-relevant educational research.

4. Research data and access in Germany

Germany has developed a comparatively sophisticated but institutionally fragmented infrastructure for research access to educational data. Over the past two decades, a network of

⁷ It is a widely used behavioural screening tool for children and adolescents, typically completed by parents, teachers, or the young people themselves. The SDQ measures five key domains: Emotional symptoms, conduct problems, hyperactivity/inattention, peer relationship problems, prosocial behaviour

accredited, specialised Research Data Centres (RDCs) has emerged under the coordination of the German Data Forum (RatSWD). These centres provide secure, legally compliant access to microdata for scientific use. In education, the main data infrastructures can be grouped into dedicated educational RDCs, official statistics with educational variables, large-scale assessment studies, higher education and student surveys, as well as specialised and linked data products. The network of specialised RDCs provide access to sensitive data on individual-level from diverse domains such as education, labour market, or pension insurance, while ensuring compliance with the General Data Protection Regulation (GDPR) and other legal frameworks on data privacy and security. For example, German microcensus data are made available through the Research Data Centre of the Federal Statistical Office in cooperation with the Research Data Centres of the Statistical Offices of the Federal States. Together, they operate as the RDCs of the Statistical Offices of the Federation and the Federal States.

Key institutions in the domain of educational research data partner within the German Network of Educational Research Data, e.g., the RDC at the Leibniz Institute for Educational Trajectories (LIfBi), is responsible for the dissemination of German National Educational Panel Study (NEPS) data, and the RDC at the Institute for Educational Quality Improvement (IQB), which manages, among others, data of the Educational Standards. Further, the RDC at the German Centre for Higher Education Research and Science Studies (DZHW) provides data access to the longitudinal student and graduate panels (tertiary education level). Access to these datasets is granted through standardised application procedures and is free of charge for researchers affiliated with academic institutions. To the paper, which focuses on longitudinal research data, particular emphasis is placed on the availability of longitudinal data sources.⁸ RDCs offer tiered access to research data, with the degree of data anonymisation increasing in proportion to the risk of re-identification. Less aggregated information is only available in a controlled, on-site environment, while downloadable Scientific Use Files (SUFs) are more anonymised. The increase in anonymisation, i.e. download, remote access and on-site access, can be found among others at the RDC of the Leibniz Institute for Educational Trajectories (LIfBi), see Skopek et al. (2016), Fuß and Wenzig (2019), and Hoffstätter and Linne (2024) for a more general introduction. Access to data typically requires a contractual agreement between the individual researcher and the RDC of the relevant institution. This agreement outlines the terms of access and use of data, and publication of research findings. In Germany's federal system, individual states also enact their state-specific laws governing data access for research purposes.

Germany's data protection framework is robust, primarily governed by the GDPR and the Federal Data Protection Act (BDSG). The GDPR aims at a uniform level of data protection throughout the European Union (EU), while the BDSG includes specific national regulations, particularly concerning the processing of personal data in the employment context. Additionally, each federal state has its own data protection laws that further regulate the

⁸ The paper therefor is not referring to cross sectional international assessment studies, such as PISA, TIMMS, PIAAC, or ICILS, to name a few, although some have longitudinal extensions, but includes few selected repeated cross-sectional datasets to indicated differences between the Scottish and the German situation.

processing of personal data by state authorities.⁹ This demonstrates that while the GDPR guarantees a common baseline emphasizing the individual rights of data owner, national and federal state laws create additional safeguards and context-specific adaptations, particularly relevant in education and research typically provided at the level of the federal states.¹⁰ Regulations on the inclusion of personal data in annual school reports therefore differ across federal states. Bavaria and Rhineland-Palatinate explicitly stipulate which categories of personal data may appear in such reports, including names, dates of birth, class affiliation, and, in some cases, information on specific school functions (Bayerisches Staatsministerium für Unterricht und Kultus, 2024a, 2024b; Ministerium für Bildung Rheinland-Pfalz, 2023). In contrast, federal states such as North Rhine-Westphalia and Berlin do not provide detailed catalogues for annual reports; instead, data processing is governed by general school data protection regulations (e.g., Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen (2024a, 2024b) and Senatsverwaltung für Bildung, Jugend und Familie Berlin (2024)), which means that the publication of names or photographs in yearbooks usually requires explicit parental consent. Furthermore, no unified institutional data access is established for this kind of data. These differences illustrate the heterogeneous frameworks for the permissible handling of student performance and personal data in educational contexts in Germany.

Nevertheless, open access to research data is widely advocated in Germany. Joint guidelines adopted by the Federal Government aim at advancing open data access by stressing the significance of free and open user access to digital scientific literature, alongside their reuse with minimal constraints. This dedication is evident in the extensive collection of open access repositories supported by German institutions, which contain a diverse array of scholarly works and research data, see Bundesministerium für Bildung und Forschung (BMBF) (2023) and Pigeot et al. (2022) for a discussion of the integration of health insurance data with epidemiological cancer registry data, highlighting both the potential and limitations of such data linkage.

Most recently, the Federal Ministry of Research, Technology and Space (BMFTR) aims to establish clear legal frameworks for data access in Germany with the Research Data Act (Forschungsdatengesetz, FDG). The proposed German Research Data Act is designed to improve and legally facilitate researchers' access to administrative and other data sources, strengthen data linkage across sectors, expand remote access through secure research environments, enhance legal certainty and data governance, and thereby promote data-driven research, innovation, and evidence-based policymaking

⁹ For instance, Section 9 of the Mecklenburg-Vorpommern Data Protection Act (DSG M-V) allows the processing of personal data for scientific research, provided certain conditions are met, see: <https://dsgvo-gesetz.de/dsg-m-v/9-dsg-m-v/>.

¹⁰ For example, the transmission of student performance data is regulated at the federal state level, leading to subtle but important differences, as in Bavaria, schools may transmit grades to supervisory authorities or the federal state statistical office where required by law, but public disclosure of grades within the classroom is considered a data transfer and is generally deemed disproportionate under data protection principles, see Bayerisches Staatsministerium für Unterricht und Kultus (2024b, Art. 85) (BayEUG Art. 85, BayLfD, 2022) and Bayerischer Landesbeauftragter für den Datenschutz (BayLfD) (2022). By contrast, in North Rhine-Westphalia the applicable school law and its implementing regulation explicitly list the categories of personal data, including performance-related information, that may be transferred between schools during a change of school, provided such transfer is necessary for the pupil's educational trajectory, see Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen (2024a) and Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen (2024b).

while safeguarding data protection and information security. Furthermore, uniform and less administratively burdensome data protection oversight for cross-state research projects is a central aspect of the draft, see BMFTR (2025) and RatSWD (2026). The latter would significantly facilitate the conduct of educational studies across federal states. In addition, plans include to establish a center for microdata at the national level.

Unlike panel surveys, datasets available in the domain of educational administrations are often cross-sectional rather than longitudinal and typically focus on the aggregate population level and do not contain individual or school level data, see Mundelius (2019). These data products include, among others, the school statistics, the vocational training statistics, and the higher education statistics.¹¹ Nevertheless, they represent the most comprehensive population-based information on education and training in Germany and are frequently used as auxiliary data – for instance, to derive population weights, to construct regional indicators, or to validate representativeness in longitudinal studies. The establishment of a national register of educational pathways (Bildungsverlaufsregister) would close an important gap regarding access to individual data from official statistics is currently under political and methodological discussion, see Artelt et al. (2019). A feasibility study conducted by the Federal Statistical Office examined potential data sources and challenges related to linking educational trajectories across the life course (Grimm et al., 2022). The long-term goal is to develop a comprehensive education module within the forthcoming population register census, capturing educational attainment and participation of individuals aged 15 and over. Such a register would substantially enhance the potential for longitudinal linkage of administrative and survey data, allowing for more continuous observation of educational careers. In summary, while official statistics currently provide predominantly cross-sectional educational data, the ongoing development of a national educational register and improved linkage infrastructures indicate a gradual convergence toward longitudinal, life-course-oriented data availability within the German research data system.

Within this ecosystem, Germany contributes to and benefits from a range of international and European large-scale longitudinal studies that provide harmonised data for cross-country comparison and trend analysis. Table 1 give a comparative overview of educational data infrastructures in Scotland and Germany, and table 2 summarises the findings of longitudinal data sources.

4.1. Adult Education Survey (AES)

Among other international studies, the Adult Education Survey (AES) is a cross-sectional survey conducted every five to six years across European countries, with data published by the statistical office of the European Union (Eurostat).¹² The AES has been conducted in Germany once every two-to-three years since 2007 as a survey on the participation and non-

¹¹ School level data may be available but would involve data requests to each of the federal states that typically involve processing fees. Although school level data is available for some federal states, no standardized data format is provided hindering further processing as required for statistical analysis.

¹² See <https://ec.europa.eu/eurostat/web/microdata/adult-education-survey>.

participation of adults in lifelong learning within the European Union.¹³ It has replaced the previous continuing education reporting system of the Federal Ministry of Education and Research (BMBF) see Bundesministerium für Bildung und Forschung (BMBF) (2024b). In accordance with standardised procedure, adults between the ages of 25 and 64 are surveyed on their participation in continuing education to gather information on the scope, content, and framework conditions of participation in formal, and informal education. The AES is responsible for the preparation of key data for education reporting in the EU member states and on the development of adult education at the European level. In the following section, we primarily focus on information regarding the German AES, embedded within the European context. Although the EU requires the AES to be conducted roughly every 6 years (e.g., 2007, 2011, 2016, 2022), some countries choose to collect data more frequently, or adjust fieldwork timelines slightly, based on national priorities or due to available resources. Hence, some countries (including Germany) conduct additional national surveys on adult education more frequently, which may overlap thematically with AES.

Since 2022 the AES target population covers individuals aged 18-69. According to the current scientific use file (Bundesministerium für Bildung und Forschung (BMBF), 2024a), the survey uses computer-assisted personal interviewing (CAPI) or computer-assisted web interviewing (CAWI), where typically a multi-staged stratified random sampling design is used, especially in countries not having a central population register available for sampling purposes. and stratified random sampling from national population registers (Microcensus sampling frame in Germany). Sample sizes vary slightly across countries and waves depending on national implementation. In the current wave, 9,820 persons are participating in the survey in year 2022. The main topics of the 2022 survey are job-related information, characteristics of the main occupation, educational background, participation in formal education (FED) in the last 12 months, participation in non-formal training in the last 12 months, informal learning or self-learning in the last 12 months, self-assessment of foreign language skills and health. Through the PIAAC Research Data Centre (RDC PIAAC), the Leibniz Institute for the Social Sciences (GESIS) offers prepared and comprehensively documented data from the latest surveys for download.¹⁴ Access to the German AES is granted through the Research Data Centre (RDC PIAAC) at GESIS, which is part of the accredited infrastructure under the German Data Forum (RatSWD). SUFs are available for research projects upon application, which must include a short project description, a data protection declaration, and a signed user agreement. In turn, data access is then provided either via remote desktop solutions or in secure on-site facilities, depending on the potential re-identification risk. While informed consent is generally obtained before the dissemination of survey data to the scientific community, there is currently no publicly available documentation or procedure indicating that linkage of AES data to other

¹³ Since the 2011/12 survey, the member states have been obliged to participate.

¹⁴ For an overview see: <https://www.gesis.org/data-on-adult-education/daten/weitere-daten-zur-erwachsenenbildung/adult-education-survey-aes>.

sources (e.g., administrative registers) is supported. Therefore, any linkage would likely require additional legal clarification and project-specific arrangements.¹⁵

4.2. German National Educational Panel Study (NEPS)

The German National Educational Panel Study (NEPS) is the largest and most comprehensive longitudinal study on educational trajectories in Germany and follows a multi-cohort sequence design, see figure 3), covering six major educational stages, which are early childhood (kindergarten and preschool), primary education (grades 1-4), lower secondary education (grades 5-9), upper secondary education (grades 10-13), post-secondary and vocational training, adult education and lifelong learning. It spans the entire educational life course, from early childhood education to professional life, providing detailed insights into how educational processes unfold over time. The NEPS started with a total of over 60,000 target persons from six cohorts, including newborns, kindergarten children, secondary school children (5th and 9th grade), first-year undergraduate students, and adults. Detailed definitions of the starting cohorts regarding the population, the survey design, and sample sizes can be found in Blossfeld and Roßbach (2019) and the data manuals provided by the NEPS.¹⁶ In the following, we provide a summary of each of the NEPS starting cohorts and table 2 condenses the details of the NEPS data.¹⁷ The NEPS aims to investigate educational trajectories, skill development, and social inequality. It provides valuable data for understanding how socio-economic background influences educational outcomes and life chances. The study is managed by the Leibniz Institute for Educational Trajectories (LIfBi) in cooperation with a German wide research network, which is responsible for data collection, curation, and dissemination. The combination of self-reported data, test-based competence measures, and linked administrative data allows researchers to examine how educational outcomes evolve over time and how contextual factors, such as family background and regional differences, shape these outcomes. The NEPS dataset includes rich longitudinal data, capturing individual educational biographies, competencies, and contextual factors. Additionally, administrative data (e.g., social security records) are linked to enhance the dataset's analytical potential.

The NEPS infrastructure is designed to ensure long-term access to high-quality educational data, and researchers can apply for data usage through the LIfBi's research data centre accredited by the German Data Forum, see figure (3). Access to NEPS data is tiered according to data sensitivity, see also figure 3. Following registration in form of a data user agreement, researchers can request Scientific Use Files (SUFs) as download files, while more detailed, less anonymised versions are available via a remote access environment, and the least anonymised data being provided onsite.

¹⁵ Based on the official documentation provided by GESIS and the RDC PIAAC, as of June 2025, no standard procedure or documentation for data linkage in the AES context is available.

¹⁶ See <https://www.neps-data.de>.

¹⁷ Further information regarding the variables and constructs contained in the NEPS is also available online in form of a structured search data base (<https://variablesearch.lifbi.de>) and an online user forum (<https://forum.lifbi.de/>).

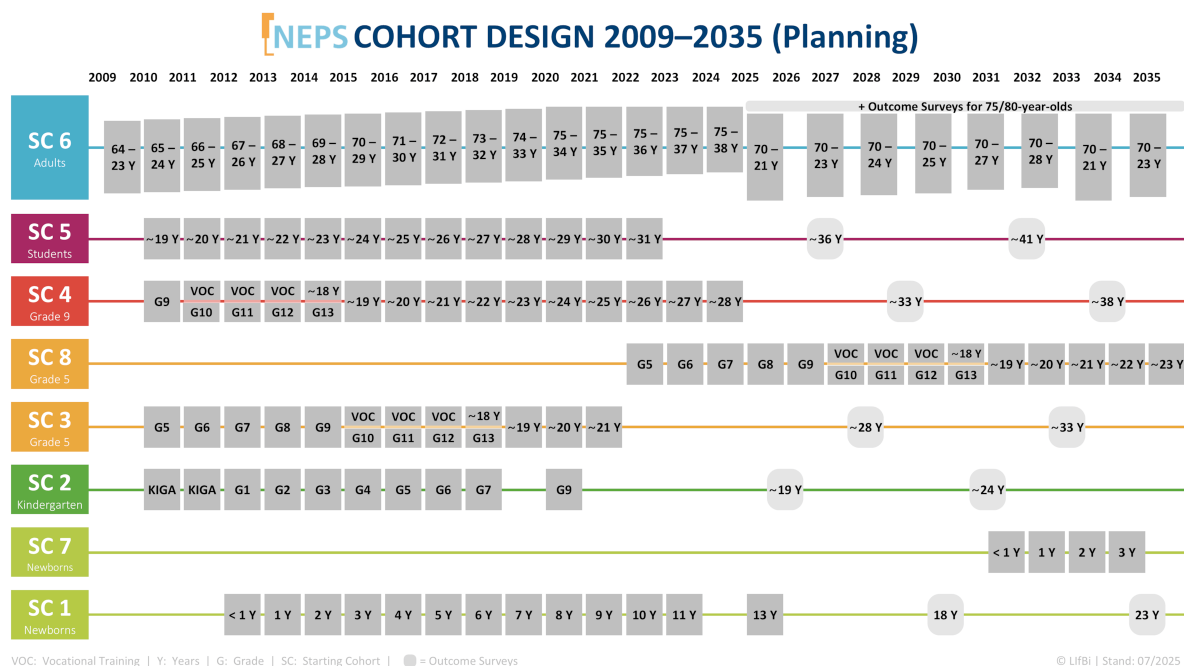


Figure 3: NEPS cohort design 2009–2035 (planning). Reproduced with permission from the Leibniz Institute for Educational Trajectories (LfBi, 2025). Status: July 2025.

4.2.1. Starting Cohort Newborns (SC1)

The study aims to create a longitudinal cohort starting with infants in their first year of life. Families are visited at home, and surveys are conducted with the children, their parents, and external childcare providers. The main research questions include the development of early skills and abilities in children and the support provided by childcare and educational settings within and outside the family, the interaction between intra-familial, and extra-familial settings, and the age at which families use childcare and education settings outside the family, influenced by child development and family background.

The sampling strategy targets all children born in Germany from February to July 2012. A two-stage disproportional stratified sampling strategy was used, with municipalities as primary units and addresses of newborns as secondary units. 84 municipalities were selected based on birth distribution, and addresses were randomly selected. The initial sample size was 8,483 addresses, resulting in 3,481 participants in the first wave. Currently the scientific use file of NEPS SC1 contains 12 survey sweeps. Further, the sampling approach of NEPS SC1 is based on the decentralized address registries provided at the level of municipalities. Currently, no central birth registry is available in Germany in contrast to Scotland. This implies that setting up cohort samples requires researchers or contracted survey institutes to communicate with a multitude of municipality registration offices to ask for addresses and adds to administrative costs of surveys.

4.2.2 Starting Cohort Kindergarten (SC2)

Starting Cohort 2 of the NEPS takes a longitudinal look at children in pre-school and elementary school, launched nationwide in winter 2010/11. It examines educational processes from kindergarten through early primary school, focusing on transitions and the development of cognitive and noncognitive skills. The study considers family background and institutional learning environments. Key research areas include the impact of early non-parental care, the role of schools in addressing skill gaps, and the interaction between home and institutional settings. It also explores educational decisions during the transition to secondary school, influenced by socio-economic status and teacher recommendations. Additional topics include migration, health outcomes, personality traits, motivation, and social behaviour. Data is collected from children, parents, educators, and school principals. The cohort follows children who attended kindergarten in 2010/11 and started school in 2012/13.

Initial surveys were conducted in kindergartens, followed by elementary schools. Sampling involved linking kindergartens to elementary schools, aiming to track children through transitions. Three groups emerged, i.e. those transitioning to NEPS schools, an augmentation sample, and children being re-contacted at home later, where a test was administered to all participants in wave 6. From wave seven onwards, all students were tested and interviewed individually, with their parents.

Also, the sampling approach for the cohort sample of NEPS SC2 had to rely on a decentralized indirect sampling approach, as no comprehensive list of all kindergarten institutions exists in Germany, also no at the level of federal states. The indirect sampling approach sampled schools and then referred to link kindergartens, see Steinhauer et al. (2015) for further details. Lack of a comprehensive sampling frame not only poses challenges to efficient sampling strategies in terms of population coverage but also limits the possibilities to conduct poststratification or the consideration weighting or calibration approaches to enhance or robust estimation.

4.2.2. Starting Cohort Grade 5 (SC3 and SC8)

Lower secondary education bridges elementary school and upper secondary or vocational paths. However, challenges remain around school choice, transitions, class repetition, and progression routes. The NEPS study Paths Through Lower Secondary School tracks students from regular and special schools starting in Grade 5, using annual surveys and competency tests. After leaving school, participants continue with individual interviews and assessments. Research focuses on competency development, educational conditions, personal outcomes, future pathways, and social integration. Data from teachers and principals include class size, student demographics, school resources, and teaching practices. Starting Cohort 3 includes 5th graders in German lower secondary schools during 2010/11. A stratified two-stage sampling method selected schools and classes. Participation was voluntary, with written parental consent required. Of 9,622 students, 5,283 participated (94.4%). Special-needs schools had a 97.1% participation rate. A refreshment sample in Wave 3 added 2,146 seventh graders, especially from Berlin and Brandenburg, to capture later transitions. Sampling followed the same two-

stage design. In 2022, NEPS introduced Starting Cohort 8, which again follows students from Grade 5 onward.

4.2.3. Starting Cohort Grade 9 (SC4)

SC4 tracks students' transitions to upper secondary school and vocational training. Data collection began in 2010/11 with group testing and individual surveys involving students, teachers, principals, and parents. The panel includes 16,425 ninth graders, with follow-up data on over 15,000 students across two waves. Students who leave school or change educational paths are individually tracked, making SC4 unique in covering vocational transitions. The cohort includes students from regular and special needs schools which are focused on learning disabilities. Sampling used a stratified multistage design, selecting schools and then classes from grades five and nine. Seven school types were defined as strata, reflecting Germany's diverse education system. Participation was voluntary, with separate consent for students and parents.

4.2.4. Starting Cohort First-Year Students (SC5)

SC5 focuses on educational trajectories and social integration during the transition to higher education. The cohort includes students enrolled in their first year at German universities and universities of applied sciences. Data collection began in winter 2010/11 and combines surveys and competency tests to examine academic success, dropout risks, and the influence of prior educational experiences.

Research questions address determinants of study choice, the role of social background, financial resources, and institutional support in shaping study progress and completion. Additional topics include employment during studies, mobility, and the development of professional competencies. Sampling targeted first-year students across disciplines using a stratified design based on institution type and field of study. Participants were recruited through cooperation with higher education institutions and student registration offices. The initial sample comprised approximately 17,000 students, with follow-up surveys conducted annually to capture changes in study conditions, aspirations, and career planning. Given the annual survey frequency, the data contains also information on the study progress of participants, however with focusing on a single cohort, the data is less suited to analyse the effects of changing institutional setting on study success as this would require repeated cohorts but focuses on the long term outcomes of higher tertiary education and the transitions into the labour market.

4.2.5. Starting Cohort Adult (SC6)

SC6 investigates lifelong learning and skill development among adults in Germany. The cohort began in 2009 and includes individuals aged 18 to 65, representing diverse educational and occupational backgrounds. The study explores how adults acquire, maintain, and apply competencies in formal, non-formal, and informal learning contexts. Key research areas include participation in continuing education, the impact of employment history, family responsibilities, and socio-economic factors on learning opportunities. It also examines

attitudes toward education, digital skills, and the relationship between learning and labour market outcomes. Sampling employed a two-stage design: municipalities were selected first, followed by random sampling of residents from population registers. The initial sample consisted of about 11,600 participants, with repeated surveys and competency assessments conducted every two-to-three year.

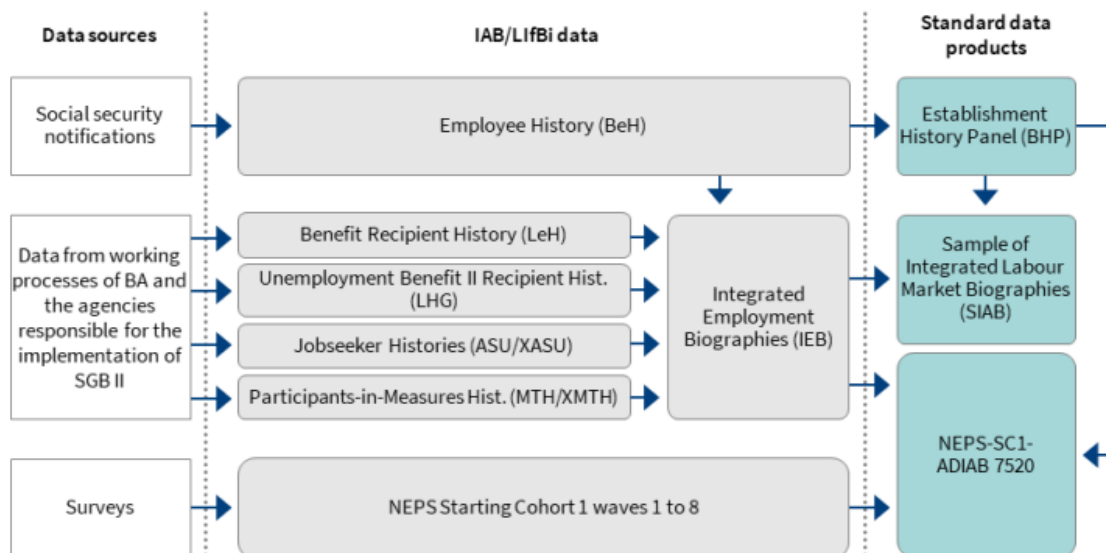


Figure 4: Data source of the NEPS-ADIAB. Extracted from Bachbauer and Wolf (2021).

4.2.6. Linking with administrative data on the labour market participation

Administrative labour-market data in Germany are provided by the Federal Employment Agency through the Research Data Centre of the Institute for Employment Research (IAB-RCD). The central data infrastructure is the Integrated Employment Biographies (IEB), which combine process-generated records on employment subject to social security contributions, marginal employment, benefit receipt under Social Code Books II and III, and registered job-seeking.¹⁸ These data cover labour-market biographies in West Germany from 1975 onwards and in East Germany from 1992 onwards (Antoni and Schnell (2019) and Müller and Möller (2019b)). They are organised as spell data with exact start and end dates, allowing employment, unemployment and benefit histories to be reconstructed at a high temporal resolution. For educational research, the most relevant linkage product is NEPS-ADIAB, jointly provided by LIfBi and the IAB, see figure 4 for an overview. NEPS-ADIAB links survey and test data from the National Educational Panel Study with administrative labour-market records from the IAB. The NEPS data provide detailed information on educational trajectories, competencies, learning environments, attitudes and family or institutional contexts, while the administrative records add long-term employment histories, benefit receipt and establishment-related information.

¹⁸ See <https://fdz.iab.de/en/our-data-products/individual-and-household-data/siab/>.

Linkage is conducted at the individual level and requires informed consent from NEPS respondents. Depending on the NEPS starting cohort, around 56 to 74 percent of respondents could be successfully enriched with administrative labour-market data (Bachbauer et al., 2022).

The combined data product substantially extends the analytical potential of NEPS. It enables researchers to study how educational experiences, competencies and transitions are related to later labour-market trajectories, including employment stability, unemployment, earnings and establishment characteristics. At the same time, the IEB do not cover all forms of labour-market participation equally. Self-employed persons, civil servants and students without relevant employment, benefit or job-seeker spells are not systematically observed. The data therefore provide highly detailed information for employment covered by the German social security and employment administration, but they do not constitute a complete register of all labour-market activity.

Access to IEB and NEPS-ADIAB is possible only under restricted conditions because of the sensitivity and density of the administrative data. Researchers need data use agreements with the relevant research data centres, and analysis is conducted through secure access modes such as on-site use, remote execution or remote desktop access. For NEPS-ADIAB, access requires agreements with both FDZ-LIfBi and IAB-FDZ and is organised through guest access at an IAB-FDZ location followed by remote processing. These access arrangements make NEPS-ADIAB an example of Germany's specialised research-data infrastructure: linkage is possible and analytically powerful, but it depends on informed consent, institutional cooperation and controlled access environments.

From a FAIR perspective, NEPS-ADIAB illustrates the role of specialised research-data infrastructures in enabling the reuse of sensitive linked data. Its analytical value does not arise solely from the richness of the NEPS survey and competence measures or the temporal depth of the IEB employment biographies, but from the institutional arrangements that make linkage, documentation and controlled access possible. Findability is supported through data catalogues and study documentation provided by the responsible research data centres. Accessibility is governed through formal data-use agreements and restricted access modes. Interoperability depends on the harmonisation of survey and administrative records, including linkage procedures, identifiers and metadata standards. Reusability is supported through persistent documentation, transparent access conditions and curated Scientific Use Files or secure analysis environments. At the same time, NEPS-ADIAB also shows that FAIRness in highly sensitive linked data does not imply open download access. Rather, the data can be FAIR under controlled access conditions, if metadata, documentation, governance procedures and reuse conditions are transparent and durable.

4.3. The network of Research Data Centres (RDCs) and emerging harmonisation via NFDI4Society

While individual linkage projects such as NEPS-ADIAB demonstrate the analytical potential of combining educational and labour-market data, recent developments in Germany have shifted

toward a more structural approach. The long-term goal is to harmonise standards for data access, linkage, and documentation across research domains in line with the FAIR principles. This task has been taken up by the National Research Data Infrastructure (NFDI) and the consortium for the social, behavioural, educational, and economic sciences (KonsortSWD-NFDI4Society), which builds upon the established network of research data centres.¹⁹

The German Data Forum (RatSWD) has accredited Research Data Centres (RDCs) since 2008, continuously expanding a network of certified data providers and fostering collaboration among them through the FDI Committee to improve and harmonize researchers' access to data. Given the strong thematic alignment between RatSWD and the National Research Data Infrastructure (NFDI) in the areas of research data management, data access and data infrastructure, RatSWD and its RDC network became the foundation of KonsortSWD in 2020. Since then, they have continued their activities within the framework of KonsortSWD, contributing to the development of a coordinated and sustainable research data ecosystem in the NFDI.

Established to secure long-term access to scientific data across disciplines, the NFDI promotes the systematic curation, documentation, and sharing of research data in line with the FAIR principles across disciplinary boundaries. Within this framework, KonsortSWD plays a central role in improving data accessibility and interoperability for education and the social sciences. KonsortSWD does not provide data itself, but coordinates and supports the existing network of accredited research data centres. It aims to harmonise access procedures, develop shared metadata standards, and strengthen interdisciplinary data linkages between surveys, administrative records, and official statistics. KonsortSWD especially develops infrastructure for the utilisation of microdata from official statistics, administrative sources and surveys. A central concern is to make it easier for researchers to access protected data and to promote innovative methods of data linkage. One central initiative is the establishment of RDCnet, a network of secure access points that allows researchers to analyse protected data at multiple locations without travelling to a single RDC.²⁰ Another example is the funding of linkage projects such as Children of Immigrants Longitudinal Survey in Four European Countries (CILS4EU) with NEPS data (CILS4NEPS), which illustrate how existing datasets can be matched to enhance analytical potential, see Dollmann, Arnold & Horr (2025). KonsortSWD also supports the long-term curation of qualitative and quantitative data collections through targeted funding, encouraging researchers to prepare their data for secondary use in accordance with FAIR principles. Sustainable funding from the German Research Foundation (DFG) ensures that the infrastructures developed will remain in place in the long-term and are continuously adapted to new technological and scientific requirements. By providing financial and technical support for documentation, anonymisation, and metadata creation, it helps data owners make previously unavailable datasets accessible via professional RDC infrastructures. Additionally, the moderated online forum Forum4MICA facilitates exchange among researchers and data providers, helping early-career scholars to navigate procedures for data use and analysis.²¹

¹⁹ See <https://www.nfdi.de> for details on the integrated and harmonising strategy.

²⁰ <https://www.konsortswd.de/en/services/rdcnet/>

²¹ <https://forum.lifbi.de/>

Through these measures, KonsortSWD contributes to the gradual harmonisation of Germany's fragmented data ecosystem, ensuring that decentralised access structures remain interoperable, transparent, and sustainable. It thus serves as a bridge between the established RDC network and emerging European data infrastructures, positioning Germany's social and educational sciences for stronger international cooperation and long-term FAIR compliance.

5. Relative advantages and potential improvements

Building on the preceding sections, this section synthesises the comparison of the relative advantages and limitations of Scottish and German research data infrastructures for longitudinal educational research. Rather than reiterating descriptive elements of individual data sources, the discussion highlights how institutional arrangements shape data availability, integration, and use. The comparison focuses on access, linkage, documentation, metadata and reuse conditions, and relates these dimensions to the FAIR principles (Wilkinson et al., 2016). We first summarise the two data ecosystems at infrastructure level and then use two illustrative design scenarios to show how institutional arrangements shape concrete longitudinal research designs.

5.1. Comparative assessment of research data infrastructure

Scottish and German research data infrastructures represent two distinct models of educational data governance. A central point of divergence lies in the organisation of data infrastructures. Scotland is characterised by a comparatively integrated system in which administrative data are managed within a centralised framework, facilitating systematic linkage across educational stages. Its strengths lie in population coverage, governed data linkage and the possibility of following individuals across institutional domains through secure research environments. Germany's system, in contrast, is more federally distributed and RDC based. While this structure supports specialised expertise, it requires additional coordination for data integration. Its strengths lie in rich longitudinal survey programmes, competence measures, large-scale assessment data and detailed documentation, while linkage across educational phases, Länder and institutional domains is less uniformly institutionalised. Further differences emerge in data governance. Although both systems operate under the GDPR, Scotland benefits from relatively unified procedures for data linkage. Germany combines federal and state-level regulation, resulting in a more heterogeneous legal environment that can increase complexity in accessing and linking data.

Table 1 summarises these differences across governance structures, major data sources and FAIR-oriented infrastructure characteristics. It distinguishes between the broader institutional architecture of data production and access on the one hand, and the FAIR-relevant conditions of findability, accessibility, interoperability and reusability on the other. This distinction is important because FAIRness is not determined by the substantive content or policy relevance of a dataset alone. Rather, it depends on the infrastructures through which data and metadata are documented, curated, accessed, linked and preserved.

Table 1: Comparative overview of educational data infrastructures in Scotland and Germany

Dimension	Scotland	Germany
Governance structure	Comparatively centralized system under the Scottish Government, with administrative education data coordinated through national agencies and infrastructure such as ScotXed, Education Analytical Services, Education Scotland, the Scottish Qualifications Authority (SQA), and the Scottish Funding Council (SFC). Sensitive and linkable data can be accessed through secure infrastructures such as National Safe Haven.	Federally distributed system in which educational data originate from Länder statistical offices, Destatis, IQB, IAB, LIfBi and other specialised data producers. Access is organised through accredited Research Data Centres and coordinated infrastructures such as RatSWD, with a strong emphasis on legal compliance, secure access and methodological documentation.
Main school-level datasets	Scottish Pupil Census, school attendance and absence data, exclusions, attainment and qualifications data, school leaver data, and participation in international assessments. The Pupil Census covers publicly funded schools and provides a central administrative basis for school-level research.	IQB-Bildungstrends, German samples from international assessments such as PISA and TIMSS, NEPS school cohorts, and official school and teacher statistics collected by the Länder and coordinated in official statistical reporting.
Adult-level and labour-market data	Scottish Household Survey (SHS), Annual Population Survey (APS), Labour Force Survey, Adult Education Survey and longitudinal resources such as Growing Up in Scotland. Linkage to wider administrative domains is possible under governed access conditions.	PIAAC, Adult Education Survey, Microcensus, NEPS adult cohorts, SOEP, and IAB administrative labour-market records. NEPS-ADIAB provides a prominent example of linking survey and competence data with employment biographies.
Higher-education data	Higher Education Statistics Agency records, Scottish Funding Council administrative data on enrolments, progression, attainment and destinations, and related UK-wide higher-education data infrastructures.	DZHW student and graduate panels, NEPS higher-education cohort, Destatis higher-education statistics, and administrative data from federal state ministries and higher-education institutions.
Findability	Supported by metadata catalogues and data portals such as the UK Data Service, the Scottish Government Open Data	Supported by catalogues and search infrastructures such as Verbund Forschungsdaten

Dimension	Scotland	Germany
Accessibility	portal and Research Data Scotland metadata services. Access ranges from open statistical outputs to controlled access for sensitive or linkable data through secure environments such as UKDS SecureLab or the National Safe Haven. Access is subject to accreditation, public-benefit review and disclosure control.	Bildung, GESIS, NFDI4Society and Research Data Centre portals. Access is typically organised through Research Data Centres, including Scientific Use Files, on-site access, remote execution or secure remote access. Procedures are usually based on data-use agreements and documented access conditions.
Interoperability	Supported by structured metadata, user guides, technical documentation, standard classifications and linkage infrastructures that enable harmonisation across administrative domains, where governance conditions allow.	Supported by extensive documentation, codebooks, sampling reports, scaling documentation, weighting information and harmonised metadata. Interoperability across domains is more challenging where identifiers, legal permissions and data standards differ across federal states or institutions.
Reusability	Reuse is facilitated by recurring administrative data collections, secure access infrastructures, metadata documentation and linkage-capable environments. Reuse remains conditional on governance approval, data protection requirements and sector-specific coverage.	Reuse is facilitated by long-term Research Data Centre infrastructures, documented Scientific Use Files, persistent metadata and secure access modes. Reusability is strong within curated survey and assessment infrastructures, but cross-domain reuse can be constrained by consent requirements, decentralised governance and heterogeneous linkage pathways.

The comparison should therefore not be understood as a simple contrast between data availability and data scarcity. Scotland provides a comparatively integrated administrative data ecosystem, whereas Germany offers a more institutionally specialised and federally distributed research-data infrastructure. The German case has considerable strengths, including the National Educational Panel Study, large-scale assessment data and well-established Research Data Centres. At the same time, the federal organisation of schooling and the division of responsibilities across Länder, domains and institutions create more dispersed arrangements for data governance, legal authorisation and linkage. In this sense, the German data ecosystem can be described as fragmented not because of a lack of high-quality educational data, but

because access to and linkage between such data are mediated by a more complex and decentralised institutional architecture.

Administrative and survey data offer complementary strengths for longitudinal research. Administrative data are generated routinely for operational and legal purposes and are particularly well suited for tracing trajectories across institutional domains because they may provide broad population coverage, regular temporal updates and reduced attrition in follow-up (Penneck, 2007). Survey data, by contrast, provide theoretically rich information on attitudes, motivations, family contexts, competencies and decision-making processes that are usually absent from administrative records. Longitudinal panel studies such as NEPS or Understanding Society are therefore central for analysing intra-individual change and life-course mechanisms (Blossfeld and Maurice, 2011). The main analytical potential lies in combining both types of data, provided that legal permissions, identifiers, metadata standards and secure access arrangements allow meaningful linkage (Künn, 2015).

Scotland's administrative model illustrates the analytical potential of integrated education data infrastructures. The Scottish Pupil Census provides recurring pupil-level information and, where governance conditions allow, can be linked to other administrative sources through secure research infrastructures (Research Data Scotland, 2025b; Scottish Government, 2025e). The Scottish model is therefore particularly strong where research questions require population coverage, consistent administrative outcome measures and longitudinal follow-up through governed linkage. However, administrative data are usually narrower in thematic scope than surveys, and individual-level coverage across all educational sectors, especially early childhood and tertiary education, remains a continuing area for development (Scottish Government, 2026).

Germany illustrates the potential and limits of a more distributed research-data ecosystem. NEPS-ADIAB provides a prominent example of survey-administrative linkage by combining NEPS survey and competence data with administrative labour-market records from the IAB, based on informed consent from respondents (Antoni and Schmucker, 2019; Bachbauer et al., 2021, 2022). The Integrated Employment Biographies provide detailed information on employment histories, benefit receipt and job-seeker records, although groups such as the self-employed, civil servants and students without relevant employment or benefit spells are not systematically covered (Antoni and Schnell, 2019; Müller and Möller, 2019b). This example demonstrates that Germany can provide highly valuable linked longitudinal data, but that such linkage is embedded in specialised infrastructures, consent procedures and controlled access arrangements.

Potential improvements therefore concern not only the creation of new datasets, but also the institutional conditions for linkage and reuse. In Germany, current reform debates point towards stronger integration of educational trajectories, for example through proposals for an educational trajectory register and persistent pupil identifiers. The RatSWD has proposed a *Bildungsverlaufsregister im Verbund* to improve the longitudinal linkage of educational trajectories across institutional phases (RatSWD, 2022). Related proposals discuss persistent

identifiers and federated Research Data Centre architectures as possible routes towards more integrated educational data infrastructures (Giar et al., 2023). Such initiatives could reduce the current dependence on project-specific linkage and improve long-term interoperability, but their implementation depends on constitutional, legal and Länder-level coordination. At the same time, both countries are moving toward greater harmonisation and improved data accessibility. Initiatives such as NFDI in Germany and continued investments in secure data environments in Scotland aim to strengthen linkage capacities while ensuring data protection.

Overall, the comparison shows that both countries provide important but differently organised infrastructures for longitudinal educational research. Scotland's main advantage lies in integrated administrative linkage and population coverage, whereas Germany's main advantage lies in rich survey measurement, competence data, methodological transparency and established Research Data Centre curation. The central challenge for Scotland is to extend thematic and sectoral coverage without weakening secure governance arrangements. The central challenge for Germany is to improve interoperability and linkage across Länder, educational phases and institutional domains while preserving strong data-protection standards. These issues are taken up in the following illustrative use cases, which apply the comparison to concrete longitudinal research designs.

5.2. Use case for comparing longitudinal data infrastructures

To make the institutional comparison more concrete, we use two illustrative design problems. These use cases are not empirical applications, but heuristic scenarios that show how national data infrastructures shape the feasibility of longitudinal educational research. They allow us to compare Scotland and Germany not only in terms of available datasets, but also in terms of the institutional conditions under which samples can be drawn, consent can be obtained, records can be linked, and data and metadata can be made findable, accessible, interoperable and reusable over time. The use cases pursue three analytical objectives. First, they illustrate how sampling frames and recruitment strategies depend on the organisation of administrative and survey infrastructures. Second, they show how consent regimes, identifier systems and linkage pathways affect the possibility of following individuals across educational and labour-market transitions. Third, they clarify that FAIRness is not simply a property of individual datasets, but is shaped by metadata standards, documentation, access governance, linkage procedures and long-term reuse arrangements.

To structure the comparison, we distinguish five analytical dimensions: (a) survey design and sample access, (b) sample size and representativeness, (c) thematic scope and measurement, (d) consent and linkage possibilities, and (e) data access, metadata and reusability. Table 2 summarises how these dimensions differ across selected Scottish and German data sources. In contrast to Table 1, which provides a broad overview of data infrastructures and FAIR-oriented access conditions, Table 2 focuses on the design-relevant characteristics that shape longitudinal research possibilities. It therefore serves as an analytical bridge between the general infrastructure comparison and the illustrative use cases below.

Table 2: Five analytical dimensions for comparing longitudinal educational data sources in Scotland and Germany

Feature class	Scotland	Germany
(a) Survey design and sample access	Recurring administrative education collections provide pupil-level information and can support longitudinal research where linkage is permitted. Key examples include the Scottish Pupil Census, school attendance and absence data, exclusions data, school-leaver destinations, and qualification/attainment data formerly provided through SQA and now through Qualifications Scotland.	German educational research relies more strongly on survey and assessment infrastructures. Key examples include NEPS as a multi-cohort longitudinal panel study, PISA, TIMSS and IQB-Bildungstrends as large-scale competence assessments, AES for adult learning, and SIAB/IEB or NEPS-ADIAB for administrative labour-market trajectories.
(b) Sample sizes and representativeness	Administrative school data such as the Pupil Census provide broad coverage of pupils in publicly funded schools and are therefore well suited for population-level monitoring and, where linkage is permitted, longitudinal follow-up. Qualification and attainment data provide recurring information on educational outcomes.	Survey and panel datasets such as NEPS and AES usually provide smaller but analytically richer samples. Representativeness depends on sampling design, response rates, weighting and panel maintenance. Administrative labour-market data such as SIAB/IEB provide broad coverage of employment subject to social security, but do not cover all labour-market groups equally.
(c) Thematic scope and measurement	Scottish administrative education data are strong for institutional indicators such as pupil characteristics, attendance, absence, exclusions, additional support needs, attainment, qualifications and school-leaver destinations. They are less flexible for measuring subjective attitudes, family processes, motivations or behavioural mechanisms.	German survey infrastructures are strong for competencies, educational trajectories, family background, psychosocial factors, migration, employment, learning environments and adult learning. NEPS is central for educational trajectories and competence development; AES captures adult learning; SIAB/IEB and NEPS-ADIAB add employment histories, unemployment spells and benefit receipt.
(d) Consent and linkage possibilities	Individual-level access and linkage are possible under controlled governance arrangements, public-benefit	Linkage in survey-based infrastructures usually depends on explicit informed consent and cooperation between several

Feature class	Scotland	Germany
	review, disclosure control and secure research environments. Relevant infrastructures include eDRIS, the National Safe Haven and Research Data Scotland/SCADR.	data holders. NEPS-ADIAB illustrates consent-based linkage between NEPS survey and competence data and IAB administrative labour-market records. Administrative datasets such as SIAB are pseudonymised and accessed under strict secure-use conditions.
(e) Data access, metadata and reusability	Access ranges from published aggregate statistics and open data products to controlled access for sensitive linked data through secure environments such as the National Safe Haven. Reuse depends on documentation, persistent metadata, stable classifications, linkage governance and disclosure-control procedures.	Access is organised through specialised Research Data Centres and secure access modes, including Scientific Use Files, on-site use, remote execution or secure remote access. Examples include FDZ-LifBi for NEPS, FDZ-IQB for competence studies, IAB-FDZ for labour-market data and GESIS infrastructures. Reuse is supported by documentation, metadata, codebooks and standardised access procedures.

5.2.1. Use case 1: A longitudinal infant cohort

To illustrate how institutional frameworks for data access influence methodological choices, we consider the planning of a longitudinal survey on the developmental environment of six-month-old infants and their mothers. Such a survey enterprise requires a carefully planned sampling strategy, explicit parental consent, and the possibility of later linkage to educational or health records. Applying the five analytical dimensions reveals fundamental structural contrasts between Scotland and Germany that directly reflect their compliance with the FAIR principles.

Survey design and sample access. In Scotland, longitudinal child studies and administrative education research can draw on high-coverage administrative infrastructure. For example, Growing up in Scotland used samples drawn from Child Benefit records, while Scottish health and education data infrastructure provide further opportunities for record-based sampling and linkage under controlled governance arrangements. In Germany, by contrast, nationally comparable survey designs often require probabilistic multi-stage sampling, frequently involving local population registers, schools or other decentralised sampling units. This reflects the absence of a single integrated national pupil register and the federal organisation of education. The difference is therefore not

simply one of efficiency versus consent, but of different institutional conditions for constructing population-based samples.

Sample size and representativeness. Scottish administrative infrastructure may support broad population coverage and longitudinal follow-up linkage, provided that legal, ethical and public-benefit requirements are met. The German survey-based design, by contrast, often provide smaller but analytically rich samples, with representativeness depending on sampling design, stratification, response rate and weighting. Their strength lies less in automatic population coverage than in detailed measurement and carefully designed survey instruments.

Thematic scope and measurement. Scotland's administrative data environment is particularly strong where research questions can be addressed using routinely collected information, for example on schooling, attendance, attainment, qualifications or transitions. However, administrative records are usually less flexible for measuring subjective attitudes, family practices or behavioural mechanisms. German survey infrastructure, such as NEPS, are stronger in this respect because they can include detailed modules on family background, educational decisions, competence development and learning environments, although this comes with higher survey costs and more limited coverage.

Consent and linkage possibilities. In Scotland, linkage can be organised through secure infrastructure such as eDRIS and the National Safe Haven, where approved researchers access pseudonymised data in controlled environments and outputs are subject to disclosure control. In Germany, linkage usually depends more strongly on explicit consent, data protection requirements and the cooperation of multiple data holders. These differences should be understood as differences in governance architecture rather than as differences in FAIR compliance. Consent arrangements shape whether and how data can be linked, but they do not themselves determine whether data are findable, accessible, interoperable or reusable.

Data access, metadata and reusability. FAIR compliance is primarily determined by the infrastructures through which data and metadata are documented, curated, accessed and preserved. In Scotland, secure access infrastructure support-controlled analysis of deidentified or pseudonymised data for approved public-benefits research. In Germany, RDCs provide documented scientific use files, access procedures and metadata that support secondary analysis. In both cases, FAIRness depends less on the substantive content of a dataset than on persistent identifiers, standardised metadata, transparent access conditions, documentation quality, interoperability of metadata standards and long-term reuse arrangements.

This illustrative case study underscores how institutional architectures shape every phase of longitudinal data production, from sampling and consent to linkage and dissemination. Scotland's model excels in integration, scale, and interoperability, whereas Germany's system offers considerable flexibility for survey-based measurement and strong traditions of explicit

informed consent and research-data curation. The comparison therefore illustrates that FAIRness is not determined by dataset content alone, but by the infrastructure through which data and metadata are documented, accessed and preserved. A sustainable European data ecosystem would need to combine linkage-capable administrative infrastructure with transparent consent procedures, rich measurement and robust access governance.

5.2.2. Use case 2: A longitudinal pupil cohort from school to work

To complement the infant example, we consider a second policy-relevant design problem: drawing a longitudinal cohort of approximately $n = 10,000$ pupils and following them for eight to ten years through the transition from school into vocational training, tertiary education, continuing education, and the labour market. Although the statistical trade-offs involved in such a design – including costs, clustering, design effects and attrition – are well known, the feasible solutions differ markedly across Scotland and Germany. These differences arise because institutional infrastructures shape three central conditions: (i) the availability and quality of sampling frames, (ii) the governance of consent and linkage, and (iii) the extent to which data and metadata can remain findable, accessible, interoperable and reusable over the study life cycle.

Survey design and sample access. In Scotland, a comparatively integrated administrative data environment can support sampling strategies that rely on register-based frames, school-level population information and stable identifiers. This facilitates transparent population definitions, enables cost-efficient sampling (for instance via size measures at the school level), and the possibility of sustained observation after pupils leave school through governed record linkage in secure research settings. In Germany, by contrast, the absence of a single national pupil register and the federal fragmentation of educational administration typically require multi-stage sampling based on school lists, federal-state-specific administrative sources or school-type stratification. Recruitment and post-school follow-up depend more strongly on active tracking, repeated contact and explicit consent procedures. These procedures may enhance transparency and participant autonomy, but they can also make systematic linkage and harmonised follow-up more difficult when identifiers, legal permissions and governance arrangements vary across the Länder or heterogeneous institutional domains. A pragmatic approach consistent with established survey practice is to sample a moderate number of schools, select a small number of classes per school, and then sample pupils within classes to reach the target size at baseline. This clustering reduces fieldwork costs but must be balanced against the risk of increased design effects and differential attrition. Critically, once respondents exit the school setting, follow-up strategies diverge i.e. systems that permit governed linkage can reduce reliance on repeated intensive tracking and mitigate selective loss to follow-up, whereas purely survey-based approaches typically require more resources to maintain contact and are more exposed to attrition biases.

Sampling size and representativeness. A baseline cohort of approximately 10,000 pupils would usually be chosen to allow robust subgroup analyses by school type, region, socio-

economic status, migration background or educational trajectories, while retaining sufficient statistical power after attrition. In Scotland, broader frame coverage and linkage-enabled follow-up may help preserve representativeness over time because key outcomes can remain observable even when survey participation fluctuates. In Germany, representativeness is more vulnerable to differential nonresponse and selective attrition once pupils leave the school context. This places greater emphasis on panel maintenance, repeated recruitment efforts, weighting adjustments and careful documentation of response processes.

Thematic scope and measurement. Both systems can support rich survey measurement of educational aspirations, skills, family resources, and decision-making at key transition points. Scotland's main advantage lies in the potential enrichment of survey data with administrative information on educational trajectories, qualifications and later transitions, which can strengthen longitudinal consistency of core outcome. Germany's survey-based infrastructures, by contrast, often provide greater flexibility for detailed modules on family context, subjective expectations, attitudes and competence development. However, sustaining comparable post-school outcome measurement at scale may be more difficult when follow-up depends primarily on repeated survey contact rather than administrative linkage.

Consent and linkage possibilities. In Scotland, governed linkage across educational and potentially other administrative domains can reduce reliance on repeated intensive tracking and may mitigate selective loss to follow-up, if access is limited to approved research purposes and secure environments. In Germany, linkage is typically more dependent on explicit informed consent, data protection requirements and cooperation between multiple data holders. This model can strengthen transparency and participant control, but it may limit linkage completeness and cross-domain harmonisation. From a FAIR perspective, the key issue is not whether one consent model is inherently more FAIR than the other. Rather, stable identifiers, harmonised metadata, transparent access conditions and durable linkage pathways are central for interoperability and long-term reusability.

Data access, metadata and reusability. In Scotland, authorised researchers may analyse pseudonymised linked data within secure research environments under accreditation, public-benefit review and disclosure control. In Germany, research data are commonly curated by specialised Research Data Centres and provided as Scientific Use Files or through secure access modes, usually accompanied by extensive documentation. Both models can support findability and accessibility through persistent metadata, catalogues and transparent access procedures. However, interoperability and reusability over long periods depend on whether identifiers, variable definitions, classifications, metadata standards and linkage pathways remain stable and sufficiently harmonised across waves, domains and institutions.

Overall, the pupil cohort example shows that FAIRness is not merely a matter of documenting a dataset after collection. It is shaped throughout the data life cycle by legal governance, identifier systems, metadata standards, linkage infrastructures and access arrangements. Scotland's model is advantageous where research depends on integrated administrative follow-up and cross-domain linkage. Germany's model is advantageous where research requires rich survey measurement, explicit consent and established research-data-centre curation. The main comparative lesson is therefore that high-quality longitudinal educational research requires both: infrastructures that enable secure linkage and long-term interoperability, and governance arrangements that ensure transparency, accountability and reusable documentation.

6. Conclusion and policy recommendations

Both Germany and Scotland face similar challenges in balancing data protection with data utility for scientific research. Institutional design, legal interpretation, and infrastructure investment decisively shape the research potential of longitudinal data systems in education and health. Even when legal and technical barriers are overcome, practical issues such as identifier harmonization, linkage error, and differential linkage rates across population groups can introduce systematic bias into longitudinal analyses, which is an issue that becomes particularly salient in cross-country comparison.

The two case studies underscore that FAIR principles and longitudinal survey methodology are jointly determined by institutional architectures. Scotland's framework tends to facilitate population coverage, linkage efficiency, and interoperability, enabling designs that are easier to plan and sustain beyond school. Germany's framework often supports richer survey measurement and stringent consent governance yet makes cross-domain linkage and long-run follow-up more resource-intensive and less uniform. A sustainable European data ecosystem would therefore benefit from combining register-enabled scalability and linkage capacity with strong consent procedures, rich measurement, and high-quality research access regimes.

This paper compared the institutional settings for data access and the availability of longitudinal data sources in empirical educational research in Scotland and Germany. Given that longitudinal data are primarily used to analyse determinants of change while controlling for initial conditions, both systems were assessed along five analytical dimensions – survey design and sample access, sample size and representativeness, thematic scope and measurement, consent and linkage possibilities, as well as data access, metadata and reusability – closely aligned with the FAIR principles of findability, accessibility, interoperability, and reusability. The analysis highlights a key structural difference between the two systems. In Scotland, data collection is established through legally mandated administrative procedures, often operating under an opt-out model when data are used for monitoring purposes. This ensures high population coverage and strong linkage capacities, effectively creating register-like data sources for cohorts between ages five and eighteen. However, this model entails a

narrower thematic focus and limited information on tertiary education. In Germany, by contrast, data collection primarily relies on survey-based designs with voluntary participation. The multi-cohort sequence design of NEPS and other national surveys increases temporal coverage but cannot fully compensate for non-response and attrition. Nevertheless, these survey data provide greater analytical depth and conceptual coherence, capturing competencies, motivations, and attitudes that are largely absent from administrative registers.

With respect to data access, both countries achieve comparable levels of timeliness and technical readiness. However, access in Germany is generally cost-free and standardized across accredited Research Data Centres (RDC), whereas researchers in Scotland may face processing charges. Such structural differences influence research equity, as cost-free access lowers barriers for early-career and publicly funded researchers. We agree with the recommendations of RatSWD (2022) and Giar et al. (2023) that the establishment of a comprehensive educational history register in Germany would represent a major step forward, especially in the view of international comparison. Furthermore, we recommend the ongoing debate on the introduction of a German law to improve the use of data for research (research data sets - *Forschungsdatengesetz*), see among others Kuhn et al. (2025) and Lehrerverband (2026). Moreover, there is the ongoing discussion in Scotland to augment the existing pupil-ID to a single-student-identifier (McKendrick, 2026), that allows researchers to better understand student trajectories through (school and vocational) education and entrance to labour market, and to evaluate the efficiency of investments in the school system. However, especially in Germany, its primary purpose should not merely be the statistical updating of register figures, but rather the traceability of individual educational trajectories across all federal states for scientific analysis. Such an infrastructure should follow the FAIR principles, include secure linkage options to labour-market and demographic registers, and ensure that consent procedures remain transparent and standardized across federal state boundaries.

Future data-policy efforts should aim to (i) strengthen harmonization of legal and technical standards between administrative and survey data, (ii) ensure long-term funding for metadata curation and documentation quality, and (iii) foster European-level cooperation through interoperable infrastructures such as the National Research Data Infrastructure (NFDI) in Germany and the UK Data Service. Ultimately, both countries can benefit from combining Scotland's strengths in population coverage and linkage efficiency with Germany's methodological depth and culture of open scientific access-thereby advancing FAIR-compliant, life-course-oriented data ecosystems for education and social research. In short, Scotland's experience suggests that the decisive bottleneck is often not analytical ambition, but the institutional capacity to produce and govern longitudinally linkable data in a way that remains reusable for science.

Acknowledgments

All authors gratefully acknowledge funding received from the Going Global Partnerships - Spring- board programme for UK-Germany ECR partnerships, project number 1269985745. Christian Aßmann acknowledges funding received from the German Research Foundation (Deutsche Forschungsgesellschaft, DFG) within priority programme 2431 (project number 539621548), as well as within NFDI4society (Project number 442494171).

References

- Antoni, M, & Schmucker, A. (2019). The research data centre of the german federal employment agency at the institute for employment research (rdc-iab): Linked microdata for labour market research. *International Journal of Population Data Science*, 4(2). <https://doi.org/10.23889/ijpds.v4i2.1141>
- Antoni, M, & Schnell, R. (2019). The past, present and future of the german record linkage center (grlc). *Jahrbücher für Nationalökonomie und Statistik*, 239(2), 319–331.
- Artelt, C, Bug, M, Kleinert, C, Maaz, K, & Runge, T. (2019). Nutzungspotenziale amtlicher Statistik in der Bildungsforschung. Ein Überblick zu Erreichtem, möglichen Chancen und anstehenden Herausforderungen. In D Fickermann & H Weishaupt (Eds.), *Bildungsforschung mit daten der amtlichen Statistik* (pp. 21–37). Waxmann Verlag GmbH. <https://doi.org/10.31244/dds.bh.2019.14.02>
- Aßmann, C, Gnambs, T, Konrad, A, & Würbach, A. (2024). *Estimation of Corona effects on mathematical competencies in grade 9 using NEPS data: Methodological companion* (LIfBi Working Paper No. 113). Leibniz Institute for Educational Trajectories. <https://doi.org/10.5157/LIfBi:WP113:1.0>
- Bachbauer, N, & Wolf, C. (2021). Neps-sc1 survey data linked to administrative data of the iab (neps-sc1-adiab) (tech. rep.). Institut für Arbeitsmarkt- und Berufsforschung (IAB), Nürnberg. <https://doi.org/10.5164/IAB.FDZD.2515.en.v1>
- Bachbauer, N, Wolf, C, & Fuß, D. (2022). Education and employment trajectories in neps-adiab: The survey data of the national educational panel study linked to administrative data of the institute for employment research. *European Sociological Review*, 38(4), 663–676. <https://doi.org/10.1093/esr/jcab057>
- Bachbauer, N, Wolf, C, Graf, T, Griebemer, S, Kaimer, S, Köhler, M, Lehnert, C, Oertel, M, & Schneider, A. (2021). Neps-sc1 survey data linked to administrative data of the iab (neps-sc1-adiab). <https://doi.org/10.5164/IAB.NEPS-SC1-ADIAB7520.de.en.v1>
- Bayerischer Landesbeauftragter für den Datenschutz (BayLfD). (2022). Bekanntgabe von Noten im Unterricht [Announcement of grades in class] [24. Tätigkeitsbericht 2022, Kapitel K.11]. Retrieved on September 24, 2025, from <https://www.datenschutz-bayern.de/tbs/tb22/k11.html>

- Bayerisches Staatsministerium für Unterricht und Kultus. (2024a). Bayerische Schulordnung (BaySchO) [Bavarian school regulation]. Retrieved September 24, 2025, from <https://www.gesetze-bayern.de/Content/Document/BaySchO>
- Bayerisches Staatsministerium für Unterricht und Kultus. (2024b). Bayerisches Gesetz über das Erziehungs- und Unterrichtswesen (BayEUG), Art. 85 - Verarbeitung personenbezogener Daten an Schulen [Bavarian Education Act, Art. 85 - Processing of personal data in schools]. Retrieved September 24, 2025, from <https://www.km.bayern.de/recht/datenschutz-an-schulen>
- Blossfeld, HP, & Maurice, Jv. (2011). Education as a lifelong process. In HP Blossfeld, HG Roßbach, & Jv Maurice (Eds.), Education as a lifelong process: The German National Educational Panel Study (neps) (pp. 19–34, Vol. 14). VS Verlag für Sozialwissenschaften.
- Blossfeld, HP, & Roßbach, HG (Eds.). (2019). Education as a Lifelong Process. The German National Educational Panel Study (NEPS). Springer VS.
- Bundesministerium für Bildung und Forschung (BMBF). (2023). Open Access in Germany - Joint Guidelines of the Federal Government and the Länder. https://www.bmbf.de/SharedDocs/Publikationen/DE/FS/772970_Open_Access_in_Deutschland_en.pdf?__blob=publicationFile&v=4
- Bundesministerium für Bildung und Forschung (BMBF). (2024a). Adult Education Survey (AES 2022 - Germany). <https://doi.org/10.4232/1.14234>
- Bundesministerium für Bildung und Forschung (BMBF). (2024b). Weiterbildungsverhalten in Deutschland 2022. Ergebnisse des Adult Education Survey. AES Trendbericht. https://www.bmbf.de/SharedDocs/Publikationen/DE/1/26667_AES-Trendbericht_2022.pdf?__blob=publicationFile&v=5
- Bundesministerium für Forschung, Technologie und Raumfahrt (BMFTR) (2025). Entwurf eines Gesetzes zur verbesserten Nutzung von Daten für die Forschung. https://www.bmftr.bund.de/SharedDocs/Downloads/DE/gesetze/forschungsdatengesetz/referentenentwurf/referentenentwurf_bessere_nutzung_von_daten_forschung.pdf
- Burns, NP, Young, D, Sherriff, A, Black, P, Blackshaw, A, & Kelly, L. (2025). Tracking the success of contextual offer students at one Scottish Higher Education Institution. Higher Education Quarterly, 79(1). <https://doi.org/10.1111/hequ.70011>
- Connelly, R, Playford, CJ, Gayle, V, & Dibben, C. (2016). The role of administrative data in the big data revolution in social science research [Special issue on Big Data in the Social Sciences]. Social Science Research, 59, 1–12. <https://doi.org/https://doi.org/10.1016/j.ssresearch.2016.04.015>
- Dollmann, J., Arnold, L., & Horr, A. (2025). CILS4NEPS – Unlocking Research Potential Through More Participants, More Schools and International Comparison: Harmonized Data for Research on Education, School-to-work Transition and Integration Processes for

- Adolescents in Germany, the Netherlands, Sweden and England. *Jahrbücher Für Nationalökonomie Und Statistik*, 245(1–2), 215–234. <https://doi.org/10.1515/jbnst-2024-0016>
- Fuchs, T, & Wößmann, L. (2007). What accounts for international differences in student performance? a re-examination using PISA data. *Empirical Economics*, 32(2-3), 433–464. <https://doi.org/0.1007/s00181-006-0087-0>
- Fuß, D, & Wenzig, K. (2019). The research data center: Making national educational panel study data available for research. In HP Blossfeld & HG Roßbach (Eds.), *Education as a lifelong process: The german national educational panel study (neps)* (pp. 361–378). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-23162-0_19
- Gao, C., McGilchrist, M., Mumtaz, S., Hall, C., Anderson, L. A., Zurowski, D., Gordon, S., Sullivan, F., & Stables, C. L. (2022). *A national network of safe havens: Scottish perspective*. *Journal of Medical Internet Research*, 24(3), e31684. <https://doi.org/10.2196/31684>
- Ghosh, S, & Kleine, L. (2024). School entry-age effect on student's affective - motivational attitudes in german elementary schools. *Early Childhood Education Journal*, 1–12. <https://doi.org/10.1007/s10643-024-01647-3>
- Giar, K, Hohlstein, F, Wipke, M, & Scharnagl, A. (2023). Konzeption eines Statistischen Bildungsver- laufsregisters in Deutschland - Entwicklungen bis 2023 und Ausgestaltungsoptionen. *WISTA - Wirtschaft und Statistik*.
- Gilbert, A, Smith, N, Knudsen, L, Jindal-Snape, D, & Bradshaw, P. (2021). Transition from primary to secondary school: Findings from the growing up in Scotland study. Scottish Government. <https://www.gov.scot/publications/transitions-primary-secondary-school/>
- Grimm, E, Herzog, O, & Rheine, S. (2022). Das Bildungsmodul des Registerzensus. *WISTA - das Wissenschaftsmagazin*, 4, 39–49. https://www.destatis.de/EN/Themes/Society-Environment/Register-census/Publications/bildungsmodul-registerzensus-042022.pdf?__blob=publicationFile&v=3
- Growing Up in Scotland. (2025). Growing up in scotland. <https://growingupinscotland.org.uk>
- Hoffstätter, U, & Linne, M. (2024). Datenzugang. Einführung in das Thema Zugang zu Daten der Sozial-, Verhaltens-, Bildungs- und Wirtschaftswissenschaften in Forschungsdatenzentren. <https://doi.org/10.5281/zenodo.13768710>
- Kuhn, A, Schleper, H, Pfafferott, M, & Zorn, D. (2025). Bildungsverlaufsregister und Schüler-ID in Deutschland. <https://doi.org/10.11586/2025054>
- Künn, S. (2015). The challenges of linking survey and administrative data. *IZA World of Labor*, 214. <https://doi.org/10.15185/izawol.214>
- Lehrerverband, PD. (2026). Schüler-ID - Chancen, Risiken und notwendige Klärungen. https://www.lehrerverband.de/wp-content/uploads/2026/01/Deutscher_Lehrerverband_Schueler_ID_Januar_2026.pdf

- Leibniz Institute for Educational Trajectories. (2025). *NEPS cohort design 2009–2035 (planning)* [Figure]. LIfBi. <https://www.neps-data.de/Project-Overview/Survey-Design>
- Matthewes, SH. (2020). Better together? heterogeneous effects of tracking on student achievement. *The Economic Journal*, 131(635), 1269–1307. <https://doi.org/10.1093/ej/ueaa106>
- McKendrick, JH. (2026). Commissioner for fair access: Annual report 2025/26. Scottish Government. Edinburgh. <https://www.gov.scot/publications/commissioner-fair-access-annual-report-2025-26/>
- Ministerium für Bildung Rheinland-Pfalz. (2023). Grundsschulordnung (GrSchulO), Übergreifende Schulordnung (ÜSchO), Föderschulordnung (FöSchulO), Berufsoberschulordnung (BBiSchulO) [School regulations of rhineland-palatinate]. Retrieved September 24, 2025, from <https://mbwwk.rlp.de/de/themen/schule/datenschutz-an-schulen>
- Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen. (2024a). Schulgesetz für das Land Nordrhein-Westfalen (SchulG NRW) [School act of north rhine-westphalia]. Retrieved September 24, 2025, from <https://bass.schule.nrw>
- Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen. (2024b). Verordnung über die zur Verarbeitung zugelassenen Daten von Schülerinnen, Schülern und Eltern (VO-DV I) [Regulation on the processing of pupil and parent data (vo-dv i)]. Retrieved September 24, 2025, from <https://bass.schule.nrw/101.htm>
- Müller, D, & Möller, J. (2019a). Giving the international scientific community access to german labor market data: A success story. In N Crato & P Paruolo (Eds.), *Data-driven policy impact evaluation* (pp. 101–117). Springer International Publishing. https://doi.org/10.1007/978-3-319-78461-8_7
- Müller, D, & Möller, J. (2019b). Giving the international scientific community access to german labor market data: A success story. *Data-Driven Policy Impact Evaluation: How Access to Microdata is Transforming Policy Design*, 101–117.
- Mundelius, M. (2019). Der Kerndatensatz auf der Basis von Individualdatenerhebungen in der Schulstatistik. Von Summendaten zu Einzeldaten. In D Fickermann & H Weishaupt (Eds.), *Bildungsforschung mit daten der amtlichen Statistik* (pp. 38–45). Waxmann Verlag GmbH. <https://doi.org/10.31244/dds.bh.2019.14.03>
- Niemann, D, Martens, K, & Teltemann, J. (2017). PISA and its consequences: Shaping education policies through international comparisons. *European Journal of Education*, 52(2), 175–183. <https://doi.org/10.1111/ejed.12220>
- OECD. (2023). *PISA 2022 Ergebnisse: Lernstände und Bildungsgerechtigkeit* (1st ed.). <https://doi.org/10.3278/6004956w>

- Office for National Statistics. (2025). Scottish Longitudinal Education Outcomes University extract in the UK. Retrieved November 17, 2025, from <https://doi.org/10.57906/ca6r-p910>
- Passaretta, G, & Skopek, J. (2021). Does schooling decrease socioeconomic inequality in early achievement? a differential exposure approach. *American Sociological Review*, 86(6), 1017–1042. <https://doi.org/10.1177/00031224211049188>
- Penneck, S. (2007). Using administrative data for statistical purposes. *Economic and Labour Market Review*, 1(10), 19–23. <http://escoe-website.s3.amazonaws.com/wp-content/uploads/2020/09/04210531/Economic-Labour-Market-Review-October-2007.pdf>
- Pigeot, I, Bongaerts, B, Eberle, A, Katalinic, A, Kieschke, J, Luttmann, S, Meyer, M, Nennecke, A, Rathmann, W, Stabenow, R, Wilsdorf-Köhler, H, Kollhorst, B, & Reinders, T. (2022). Verknüpfung von Abrechnungsdaten gesetzlicher Krankenkassen mit Daten epidemiologischer Krebsregister: länderspezifische Möglichkeiten und Limitationen. *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz*, 65(5), 615–623. <https://doi.org/10.1007/s00103-021-03475-x>
- RatSWD. (2022). Positionspapier des RatSWD: Aufbau eines Bildungsverlaufsregisters: Datenschutzkonform und forschungsfreundlich. Rat für Sozial- und Wirtschaftsdaten (RatSWD). <https://www.konsortswd.de/wp-content/uploads/Positionspapier-RatSWD-Aufbau-eines-Bildungsverlaufsregisters.pdf>
- RatSWD. (2025). Positionspapier des RatSWD: Datensilos aufbrechen - Forschung braucht gute Daten! Rat für Sozial- und Wirtschaftsdaten (RatSWD). <https://www.konsortswd.de/wp-content/uploads/RatSWD-Forderungen-21-LP.pdf>
- RatSWD. (2026). Stellungnahme des RatSWD zum Referentenentwurf eines Gesetzes zur verbesserten Nutzung von Daten für die Forschung (Forschungsdatengesetz). Rat für Sozial- und Wirtschaftsdaten (RatSWD). https://www.konsortswd.de/wp-content/uploads/RatSWD_Stellungnahme_Forschungsdatengesetz_Verbaendebeteiligung_Webseite.pdf
- Research Data Scotland. (2025a). *Using the Researcher Access Service: A guide for applicants*. Research Data Scotland. <https://www.researchdata.scot/media/mmjfhomy/using-the-researcher-access-service-a-guide-for-applicants-september-2025.pdf>
- Research Data Scotland. (2025b). Linkable pisa dataset available for research. Research data Scotland. Retrieved February 13, 2026, from <https://www.researchdata.scot/news-and-insights/linkable-pisa-dataset-available-for-research/>
- Research Data Scotland. (2025c). Research data scotland. <https://www.researchdata.scot/>
- Research Data Scotland. (2025d). Researcher Access Service [Online; accessed 26-March-2025]. <https://www.researchdata.scot>

- Research Data Scotland. (n.d.). *Short guide to eDRIS*. Research Data Scotland.
<https://www.researchdata.scot/media/olujaghh/edris-short-guide.pdf>
- Sawyer, G, Fraser, A, Lawlor, DA, Sharp, GC, & Howe, LD. (2025). Associations of adolescent menstrual symptoms with school absences and educational attainment: Analysis of a prospective cohort study. *NPJ Sci. Learn.*, 10(1), 54. <https://doi.org/10.1038/s41539-025-00338-x>
- Scottish Government. (2023). *Health and social care: data strategy*. Scottish Government.
<https://www.gov.scot/publications/data-strategy-health-social-care-2/pages/12/>
- Scottish Government. (2025a). Independent schools in Scotland: register.
<https://www.gov.scot/publications/independent-schools-in-scotland-register/>
- Scottish Government. (2025b). Longitudinal Education Outcomes (LEO) from Universities: 2022-23: Scotland. Retrieved October 29, 2025, from
<https://www.gov.scot/publications/longitudinal-education-outcomes-leo-from-universities-2022-23-scotland/>
- Scottish Government. (2025c). School contact details.
<https://www.gov.scot/publications/school-contact-details/>
- Scottish Government. (2025d). School level summary statistics.
<https://www.gov.scot/publications/school-level-summary-statistics/>
- Scottish Government. (2025e). Scottish Exchange of Data: school-pupil census. Retrieved May 6, 2025, from <https://www.gov.scot/publications/scottish-exchange-of-data-school-pupil-census/>
- Scottish Government. (2025f). Scottish Government statistics. Retrieved May 6, 2025, from <https://www.gov.scot/collections/scottish-government-statistics/#education>
- Scottish Government. (2026). Progress as a precursor to a pivot: Fair access in Scotland in 2026 and beyond annual report 2025/26.
- Scottish Qualification Authority. (2025). Freedom of Information disclosure log.
<https://www.sqa.org.uk/sqa/93617.html>
- Senatsverwaltung für Bildung, Jugend und Familie Berlin. (2024). Schulgesetz für das Land Berlin (SchulG Bln) [School act of berlin]. Retrieved September 24, 2025, from <https://gesetze.berlin.de/bsbe/document/jlr-SchulGBErahmen>
- Sixt, M, & Aßmann, C. (2020). The influence of regional school infrastructure and labor market conditions on the transition process to secondary schooling in germany [Publisher: Waxmann]. *Journal for Educational Research Online*, 12(2), 36–66.
<https://doi.org/10.25656/01:20972>
- Skopek, J, Koberg, T, & Blossfeld, HP. (2016). RemoteNEPS - An Innovative Research Environment. In HP Blossfeld, J von Maurice, M Bayer, & J Skopek (Eds.),

- Methodological issues of longitudinal surveys: The example of the national educational panel study (pp. 611–626). Springer Fachmedien Wiesbaden.
https://doi.org/10.1007/978-3-658-11994-2_34
- Smyth, E, & Privalko, I. (2023). The long road to secondary school: Background, home learning environment, and transition difficulties in scotland. *Research Papers in Education*, 38(6), 847–864. <https://doi.org/10.1080/02671522.2022.2065520>
- Stanat, P, Schipolowski, S, Schneider, R, Weirich, S, Henschel, S, Sachse, KA, & Lohbeck, A. (2023). IQB-Bildungstrend 2022: sprachliche Kompetenzen am Ende der 9. Jahrgangsstufe im dritten Ländervergleich.
- Steinhauer, HW, Aßmann, C, Zinn, S, Goßmann, S, & Rässler, S. (2015). Sampling and weighting cohort samples in institutional contexts: The national educational panel study cohort samples of kindergarten children, students in grade 5 and in grade 9. *AStA Wirtschafts- und Sozialstatistisches Archiv*, 9(2), 131–157.
<https://doi.org/10.1007/s11943-015-0162-0>
- UK Government. (1978). National health service (scotland) act. Retrieved July 14, 2023, from <https://www.legislation.gov.uk/ukpga/1978/29/contents>
- UK Government. (1980). Education (scotland) act. Retrieved July 14, 2023, from <https://www.legislation.gov.uk/ukpga/1980/44/contents>
- UK Government. (2025). Open Government License for public sector information.
- University of Bristol. (2025). Avon longitudinal study of parents and children.
<https://www.bristol.ac.uk/alspac/>
- Volodina, A, Weinert, S, Washbrook, E, Waldfogel, J, Kwon, SJ, Wang, Y, & Perinetti Casoni, V. (2023). Explaining gaps by parental education in children’s early language and social outcomes at age 3–4 years: Evidence from harmonised data from three countries. *Curr. Psychol.*, 42(30), 26398–26417. <https://doi.org/10.1007/s12144-022-03754-z>.
- Wilkinson, MD, Dumontier, M, Aalbersberg, IJ, Appleton, G, Axton, M, Baak, A, Blomberg, N, Boiten, JW, Da Silva Santos, LB, Bourne, PE, Bouwman, J, Brookes, AJ, Clark, T, Crosas, M, Dillo, I, Dumon, O, Edmunds, S, Evelo, CT, Finkers, R, ... Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3(1), 160018. <https://doi.org/10.1038/sdata.2016.18>.
- Zinn, S, Würbach, A, Steinhauer, HW, & Hammon, A. (2020). Attrition and selectivity of the NEPS starting cohorts: An overview of the past 8 years. *AStA Wirtschafts- und Sozialstatistisches Archiv*, 14(2), 163–206. <https://doi.org/10.1007/s11943-020-00268-7>.

List of Acronyms

ADR	Administrative Data Research
AES	Adult Education Survey
ALSPAC	Avon Longitudinal Study of Parents and Children
BDSG	Federal Data Protection Act (Bundesdatenschutzgesetz)
BMBF	Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung)
CAPI	Computer-assisted personal interviewing
CAWI	Computer-assisted web interviewing
CHI	Community Health Index
CILS4EU	Children of Immigrants Longitudinal Survey in Four European Countries
DFG	German Research Foundation (Deutsche Forschungsgemeinschaft)
DZHW	German Centre for Higher Education Research and Science Studies
eDRIS	Electronic Data Research and Innovation Service
Eurostat	Statistical Office of the European Union
FAIR	(F)indability, (A)ccessibility, (I)nteroperability, and (R)eusability
FDZ	Forschungsdatenzentrum (Research Data Centre)
FED	Formal education
FOI	Freedom of Information
FORUM4MICA	Making Information Commonly Available; online forum on research data
GDPR	General Data Protection Regulation
GESIS	Leibniz Institute for the Social Sciences
GUS	Growing Up in Scotland
IAB	Institute for Employment Research
IEB	Integrated Employment Biographies
ICILS	International Computer and Information Literacy Study
IQB	Institute for Educational Quality Improvement
LifBi	Leibniz Institute for Educational Trajectories
NEPS	National Educational Panel Study
NEPS-ADIAB	National Educational Panel Study linked to administrative data of the Institute for Employment Research
NFDI	National Research Data Infrastructure (Nationale Forschungsdateninfrastruktur)
NFDI4Society	NFDI4Society Consortium for the social, educational, behavioural and economic sciences within the National Research Data Infrastructure
NHS	National Health Service
OECD	Organisation for Economic Co-operation and Development
ONS	Office for National Statistics
PBPP	Public Benefit and Privacy Panel
PHS	Public Health Scotland
PIAAC	Programme for the International Assessment of Adult Competencies
PISA	Programme for International Student Assessment
RAS	Research Access Service
RatSWD	German Data Forum (Rat für Sozial- und Wirtschaftsdaten)
RDCnet	Secure data access point network of Research Data Centres
RDCs	Research Data Centres
SC	Starting Cohort

SDQ	Strengths and Difficulties Questionnaire
SUFs	Scientific Use Files
TIMSS	Trends in International Mathematics and Science Study
TREs	Trusted Research Environments
VPN	Virtual Private Network

Imprint:

Contact:

Leibniz Institute for Educational Trajectories (LifBi)

Wilhelmsplatz 3

96047 Bamberg

christian.assmann@lifbi.de

<https://www.lifbi.de>

KonsortSWD Working Paper:

KonsortSWD – NFDI4Society, as part of the National Research Data Infrastructure, is expanding offerings to support research with data in the social, behavioural, educational, and economic sciences. Our mission is to strengthen, expand and deepen the research data infrastructure for the study of society. It should be user-oriented and consider the needs of the research communities. An important cornerstone is the network of research data centres that has been built up for more than two decades by the German Data Forum (RatSWD). This series contains articles on research data management that are produced in the context of KonsortSWD. Articles that have been externally and double-blind peer-reviewed are marked accordingly.

KonsortSWD is funded by the German Research Foundation (DFG) within the framework of the NFDI – project number: 442494171.

Bamberg, July 2026



This publication is licensed under a Creative Commons Attribution 4.0 (CC-BY_4.0):

<https://creativecommons.org/licenses/by/4.0/>

DOI: 10.5281/zenodo.21369783

Citation suggestion:

Aßmann, C., Bergrab, M., Burns, N., Csoban, T., Ghosh, S., Kelly, L., Kirkpatrick, R., Stewart, R., Schlücker, F., & Stingl, D. (2026). *Scottish and German Perspectives on Access to Longitudinal Research Data in Education*. KonsortSWD Working Paper 18/2026. Konsortium für die Sozial-, Verhaltens-, Bildungs- und Wirtschaftswissenschaften (KonsortSWD – NFDI4Society).

<https://doi.org/10.5281/zenodo.20431657>