



Higher Education Authority

A Data Plan for Equity of Access to Higher Education

*Trutz Haase – Social & Economic Consultant
17 Templeogue Road, Terenure, Dublin 6W
Tel. 01 490 8800 Email: mail@trutzhaase.eu*

In collaboration with

*Jonathan Pratschke – University of Salerno,
Fisciano, 84084 - Salerno, Italy
Email: jpratschke@unisa.it*

December 2017



Table of Contents

1.	Executive Summary	3
2.	Introduction.....	8
3.	Research Team	10
4.	A New Data Plan for Monitoring Equity of Access to Higher Education	11
5.	Background.....	17
6.	Logic Model	22
7.	Assessment of existing approach to measuring equity of access	25
8.	Small-area deprivation indices as a proxy for social class	30
9.	Measuring other attributes: students with a disability, Irish Travellers	33
10.	Other target groups: mature students, part-time students, further education award holders	34
11.	Dealing with international flows	37
12.	Coverage of key datasets	38
13.	Developing a Shared Research Agenda	43
14.	Change Management	46
15.	Conclusions.....	49
16.	Bibliography	51

1. Executive Summary

This data plan was developed by Trutz Haase and Jonathan Pratschke with the aim of improving the capacity of the HEA to measure and monitor equity of access. The plan exploits the possibilities created by new technologies and administrative databases and, if implemented, would greatly expand the scope for research on educational inequalities in Ireland. It relies on linking and geocoding datasets, whereby information from different sources is brought together either at the level of the individual or for Small Areas of residence. In this way, it is possible to overcome the difficulties posed by selective non-response, manual coding of occupations, discordant measures of socio-economic group, incremental inequalities, changes in family structure and the composition of broad occupational categories, which have hampered research on equity of access to higher education until now.

Since their introduction in 2007, the current “Equal Access” data collection procedures have played a valuable role in guiding policies and practices. Although these procedures have drawbacks – which partly motivate the present study – they should arguably not be abandoned until new techniques for assessing equity of access have been introduced and evaluated. Allowing for a period of transition characterised by an overlap between the two estimation approaches would maximise the opportunity for learning and help to minimise the risks associated with the introduction of a new framework. The approach outlined in this data plan should also be carefully assessed following implementation. This report shows that there are good reasons for replacing the Equal Access Survey as the main source of data on equalities in access to higher education. Although there is no ‘gold standard’ against which these different approaches can be compared, we are confident that the techniques outlined in this report can improve on existing methods for measuring and monitoring inter-group disparities in educational participation.

The new data plan is based on the idea of estimating the socio-economic position of students by identifying the Small Area in which their parents reside. This way of identifying socio-economic position – using an aggregate-level indicator as proxy – provides an accurate and effective measure of social background. Although some individuals may be misclassified (for example, due to variability within Small Areas), these discrepancies will tend to cancel out as data are aggregated to form larger categories. If there are systematic trends, these can also be identified, so the final estimates will be much more precise than those which are currently used to inform policies in this sector.

Aggregate Small Area data are widely used as a proxy for individual socio-economic position at the international level – particularly in the field of health – and this approach has also been adopted by the Department of Education and Skills to improve the targeting of social inclusion interventions such as DEIS. By identifying and geocoding the students who complete post-primary education, who apply to higher education institutions, who take up a place and who successfully complete their course of study, the information needed for monitoring equity of access can be obtained without collecting new data. Other groups – including students with a disability, Travellers or mature students – can be identified, maximising the ‘value-for-money’ provided by existing databases and reducing the “respondent burden” on students as well as Higher Education Institutions.

Not only is it increasingly onerous to collect questionnaire data, it is also increasingly problematic due to selective non-response, differences in how the job of the head of household is described, problems with occupational coding, social mobility, demographic transformations and associated difficulties with interpreting results. Alternative approaches based on linked administrative datasets are being adopted with greater frequency in Ireland and the UK, precisely because they allow us to answer more questions, to gain greater insights and to obtain more accurate estimates. Although they pose some challenges, they are increasingly feasible in technical terms due to the development of powerful datasets containing information on service users across different sectors.

We show in this report that existing measures of socio-economic group based on occupational classifications can yield misleading estimates of inequalities in access to higher education due to the factors listed above. Because the HP Deprivation Index captures, above all, the relativities between residential areas in terms of their population attributes, it is well-placed to overcome these difficulties, making it an obvious choice as a proxy measure of socio-economic position. Rather than measuring the relative size of the Higher Professional, Non-Manual and Semi-Skilled Manual social classes, which are only roughly ordered from highest to lowest, we can analyse a continuous distribution of scores. We do not have to assume that all deprived individuals are clustered within uniformly-deprived areas, as we can use deprivation scores to estimate the extent of deprivation within each Small Area. As noted before, this yields accurate estimates of inequalities after we have aggregated to broad categories of affluence/deprivation.

In recent decades, research in Ireland and other economically-advanced countries has demonstrated the existence of significant disparities in educational participation and attainments according to socio-economic position (however this is measured). When monitoring equity of access to higher education, it is important to situate these disparities in relation to the education system as a whole. Social inequalities accumulate across the life cycle and develop incrementally within each level of the education system. “Joined-up policies” which seek to understand how inequalities are structured at a systemic level require a “joined-up evidence base”. As far as higher education is concerned, this means analysing the interface between post-primary, further and higher education and seeking to understand how participation and achievement are structured in and across the different components of the education system. A crucial and immediate benefit of the new data plan is that it will provide reliable information on how the overall inequalities observed in relation to participation in higher education are generated, including school experience, examination results and decisions regarding the timing of labour market entry.

It is helpful to think of the new data plan as contributing to the national data infrastructure, where this is defined as a relatively stable, shared information resource which can serve several purposes and is of strategic importance. This enables us to appreciate how the data plan relates to other aspects of the national data infrastructure, such as the system of PPSN codes, addresses and Eircodes, Ordnance Survey maps, Census Small Areas and the Census of Population itself. This infrastructure comprises a number of different elements, ranging from supporting legislation to data protocols, public and private organisations, new and existing datasets and statistical definitions. This data plan is about bringing these different elements into a new configuration in order to provide policy-makers and administrators with reliable and accurate information on educational equalities. As noted above, the immediate advantages generated by the data plan extend to the post-primary education system and

include the possibility of monitoring public interventions to promote educational participation (involving the DEIS programme, for example).

When viewed from this perspective, the data plan necessarily touches on the activities, goals and interests of several different organisations, both public and private. Its implementation represents a collective endeavour and requires the active participation of several stakeholders, regardless of whether they are data-holders. The data plan is founded on two key databases, the DES Post-Primary Online Database (which contains a comprehensive register of post-primary school students in Ireland) and the HEA Student Record System (which contains a comprehensive register of students at publicly-funded Higher Education Institutions in Ireland). It is proposed to link these datasets at the individual level using PPSN numbers, validated using names, addresses and Eircodes. The family address of students in both datasets will most likely have to be geocoded separately, with the aim of linking them to a Census Small Area and associating them with up-to-date census data at this fine level of spatial definition.

P-POD contains information on country of birth, ethnicity and nationality, facilitating the identification of migration background and Travellers. Although the quality and coverage of these fields may not be fully satisfactory at the moment, these aspects of the database could be improved in the future. The DES can match P-POD with information on students who sat the Junior and Leaving Certificates, providing additional information on educational progress. Data on examination results could be provided directly by the State Examination Commission, although this information is already available for most higher education students via the SRS. As far as disabilities are concerned, it would be possible to identify students in receipt of support under the Fund for Students with Disabilities (FSD); the HEA plans to integrate this information into the SRS (from 2019). The census would provide the crucial proxy measure of socio-economic position (in the form of the Pobal HP Small Area Deprivation Index).

This combined dataset could, in the future, be extended by adding individual-level data from UCAS, with a view to identifying successful applicants to higher education institutions in Northern Ireland and Great Britain (which are not currently taken into account). Further data could come from the CAO and SOLAS, providing better control over entrance routes and trajectories. An important gap in terms of current data availability relates to private Higher Education Institutions, which do not provide data for inclusion in the HEA's Student Record System and do not have an alternative, centralised database. This gives rise to a structural gap in coverage which – together with the flow of students to universities in the UK – could lead to bias when estimating equity of access to higher education. A strong case could therefore be made for instituting a statutory obligation for all higher education institutions to provide individual-level data for inclusion in the SRS.

The direct exchange of datasets between organisations may pose difficulties regarding data protection, security and access. The data plan thus foresees a role for a “trusted third party”, effectively centralising the process of data linking, anonymisation and providing secure access for research and analysis.

Deferred entry to higher education as a mature student can be modelled across age groups in order to reduce the risk of confounding. The propensity to return to education as a mature student evolves under the influence of various factors, such as the economic cycle, cultural trends and government policy and provides a means of addressing educational

inequalities. Part-time students and (more generally) those on 'flexible learning' programmes can already be identified within the Student Records System. As in the case of mature students, these groups cannot be identified independently within the population and the individuals concerned cannot be targeted before they actually enrol. It is appropriate to encourage mature, part-time and flexible students, as these forms of educational participation represent pathways for enrolment that may be attractive to members of disadvantaged social groups, as they provide ways of reconciling participation in higher education with either paid employment or caring responsibilities. Another target group involves people who make the transition from further to higher education. This is also a potentially important pathway for enabling students from disadvantaged and under-represented social groups to enrol in higher education, and these students could potentially be identified using data from the new Programme Learner Support System (PLSS) database, which is held by SOLAS.

As the data plan relies on data sources which have not previously been used for monitoring, a change management plan is included in order to provide a 'road map' for completing the transition to the new assessment framework. The change management plan describes the main steps in the implementation of the new framework and specifies the key stakeholders and agencies involved as well as the inputs, outputs and resources required. It is not possible to specify all tasks and their precise timing in advance, as this will depend on factors which are not under the control of the HEA. We therefore indicate the key steps (see Tables 1 and 2) that must be undertaken if the data plan is to be implemented, together with a rough indication of the time required and the organisations involved. As agreement is reached with stakeholders, and the main elements of the data plan are defined, a more detailed timeframe can be established.

Table 1 Overview of implementation schedule: short-term goals

<i>Phases and steps</i>	<i>Stakeholders involved</i>	<i>Expected outputs</i>	<i>Suggested timeframe</i>
Phase 1			
Achieve agreement among stakeholders; develop a collective understanding of data plan	HEA, DES, SEC, CAO, CDETB, SOLAS, CSO	Agreement, in principle, among stakeholders regarding procedures and purposes of data-sharing	Within 6 months
Phase 2			
Improve collection of address information (including Eircodes) for families and PPS numbers for students	HEA, DES, CAO, SOLAS	Revise data definitions and collection procedures; commission data protection impact assessment; identify requirements in terms of 'enabling legislation'	Within 2 years
Phase 3			
Identify an appropriate 'trusted third party' and agree terms for sharing and linking data	HEA, DES, CSO	Formal agreements between the HEA, DES, CSO and any other organisations that are willing to participate	Within 2 years

Building on Phases 1-3, the medium-term goals shown in Table 2 can be identified. It is difficult to establish a timeframe for these aspects of the plan, as they will depend on the outcome of earlier phases.

Table 2 Overview of implementation schedule: medium-term goals

<i>Phases and steps</i>	<i>Stakeholders involved</i>	<i>Expected outputs</i>
Phase 4		
Geocode addresses and verify PPS numbers; link individuals to Small Area by address of parents	HEA, DES	Linked dataset containing aggregate Small Area data derived from SRS and P-POD
Phase 5		
Join P-POD and SRS data using PPS numbers	HEA, DES, CSO	Linked dataset containing individual-level data based on agreed specifications
Integrate CAO data on applicants and SOLAS data on applicants from FE and (if possible) Small Area data from UCAS on new entrants	HEA, DES, CAO, SOLAS, UCAS, CDETB, CSO	More accurate estimates controlling for international flows, transitions from FE and other characteristics
Phase 6		
Calculate participation rates for post-primary students, Leaving Certificate students and those with specified exam results	HEA	Estimates of participation rates for higher education and analysis of differentials involving target groups

2. Introduction

In this report, we set out a data plan that is in line with the goals and objectives specified in the National Plan for Equity of Access to Higher Education 2015-2019, which stipulates that the student body entering, participating in and completing higher education at all levels should reflect the diversity and social mix of Ireland's population. This implies that measures to encourage and support under-represented groups are needed and that the data plan should focus on monitoring differentials in relation to these. Objective 3.2 of the National Plan also indicates that it is necessary "[t]o review current and new data to see how this may be developed to identify geographic areas with high levels of disadvantage; and to analyse rates of participation in higher education from those areas"¹.

This data plan seeks to overcome the weaknesses inherent in current approaches to monitoring equity of access and exploits the possibilities created by new technologies and administrative databases. It relies on linking datasets, whereby information from different sources is brought together, either at the level of the individual or the Small Area of residence. This raises complex issues regarding privacy, anonymity and data security and matching individual-level data would most likely require an explicit legislative mandate and the involvement of a "trusted third party" to match datasets whilst maintaining anonymity.

As the data plan relies on data sources that have not previously been used for monitoring, a change management plan is included in order to provide a 'road map' for completing the transition to the new assessment framework. The change management plan describes each step in the implementation of the new framework and specifies the key stakeholders and agencies involved and the inputs, outputs and resources required.

Since their introduction in 2007, the current Equal Access data collection procedures have played a valuable role in guiding policies and practices. Although these procedures have some drawbacks – which partly motivate the present study – they should not be abandoned until new techniques for assessing equity of access have been evaluated. Allowing for a period of transition characterised by an overlap between the two estimation approaches will maximise the opportunity for learning and help to minimise the risks associated with the introduction of a new framework.

This data plan is centred on the possibility of estimating the socio-economic position of students by identifying the Small Area in which their parents reside. This way of identifying socio-economic position, using an aggregate-level indicator as proxy, provides a reliable and effective measure of social background. It is widely used at the international level – particularly in the field of health – and has been adopted in recent years by the Department of Education and Skills in order to improve the targeting of social inclusion interventions in relation to school students in Ireland. This is an important precedent and (as we will show below) provides a crucial source of baseline data for identifying inequalities in access to higher education.

¹ HEA (2015a) *National Plan for Equity of Access to Higher Education 2015-19*. Dublin: Higher Education Authority, p. 28.

The idea at the core of this project is that by identifying and geocoding the students who complete post-primary education, who apply to higher education institutions, who take up a place and who successfully complete their course of study, the information needed for monitoring equity of access can be obtained without collecting new data. Other groups, like students with a disability or mature students, can also be identified, maximising the ‘value-for-money’ provided by existing databases and reducing the burden on Higher Education Institutions (HEIs). This is in harmony with EU directives regarding data sharing and cross-sectoral collaboration. It can sustain a flexible and ambitious research agenda which interrogates the ‘upstream’ and ‘downstream’ dynamics of educational inequalities and provides policy-makers with reliable and up-to-date information. This raises a number of complex issues including how to define and measure social background, account for distinct educational pathways, assess incremental inequalities and identify individuals, whilst ensuring data security. All of these issues are addressed in detail in this report.

Data linkage across administrative datasets relating to post-primary and third-level education holds great promise for evidence-based policy-making. The opportunity to conduct policy-relevant research and to tackle more complex research questions would expand enormously. Many Government Departments, statutory bodies and agencies are already engaged in an analogous process, and there is an inexorable trend towards a more innovative and integrated use of existing data sources. It is increasingly necessary for *all* Departments to work at the frontier of data access and data linking in order to obtain the information that they need for effective policy development, delivery and evaluation.

This data plan was developed under the guidance of a Steering Group at the HEA, and the authors would like to thank the members of this group for their helpful comments and observations. We are particularly grateful to Professor Pat Clancy for his stimulating and challenging observations and for sharing his insights into the higher education system in Ireland. We are also very grateful to participants from stakeholder organisations who took the time to engage with this project and to provide information on existing datasets, data collection procedures and definitions. As is typically the case, we take full responsibility for any inaccuracies or errors in the report, although we have gone to considerable lengths to consult with stakeholders and to check all information we received on the various datasets.

3. Research Team

Trutz Haase has been working as an independent research consultant for more than 20 years and has extensive experience of completing studies, assessments and data plans on themes such as equity of access. He has particular expertise in the design and execution of evaluation studies using existing and new data sources and in the development of frameworks for resource allocation, monitoring and impact assessment. His consultancy work focuses on social inclusion but cuts across several fields, including early years education, child development, family well-being, health, neighbourhood processes and local and regional development.

Trutz Haase is well-known amongst administrators, policy-makers and researchers in Ireland for his work on the measurement and analysis of deprivation, including ways of improving the targeting of health and social services by taking advantage of geocoding techniques and data linkage using residential address. In recent years, several Government Departments and Agencies have adopted solutions that hinge upon the use of the Haase-Pratschke Deprivation Index in order to improve the targeting and efficacy of their interventions and programmes. This typically involves geocoding client addresses using large administrative databases and then linking them to aggregate-level measures of socio-economic position via the Small Area of residence. When operationalised within an appropriate analytical framework, with the required control variables and appropriate functional specifications, these indicators yield reliable estimates of inequalities and impacts and facilitate the development of integrated solutions for programme management.

Jonathan Pratschke is Assistant Professor of Economic Sociology at the University of Salerno and has collaborated with Trutz Haase for almost two decades. An Irish citizen, Prof. Pratschke is an expert in research design and the specification and estimation of statistical models. He is a highly-qualified methodologist with considerable experience in the estimation of hierarchical, longitudinal and multi-group Structural Equation Models. He pioneered the application of Confirmatory Factor Analysis in the measurement of deprivation at aggregate level. He is particularly interested in identifying innovative techniques for refining and testing scientific theories and hypotheses regarding the social mechanisms underlying inequalities in the fields of health, education, well-being and social cohesion. He has participated in a number of national and international research projects, including projects funded by the Italian Government² and by the European Commission³.

² Including the national project entitled “Family background, beliefs about education and participation in Higher Education: An experiment integrated with a longitudinal survey”, coordinated by Prof. Antonio Schizzerotto and funded by the Italian Government in 2012.

³ Including the European project entitled “Social Polis: Social Platform on Cities and Social Cohesion” (2008-2010) and the project on “Inequalities and Multiple Discrimination in Access to Health”, funded by the Fundamental Rights Agency of the European Union (2010-2013).

4. A New Data Plan for Monitoring Equity of Access to Higher Education

Equality in social opportunities is a fundamental, high-level policy commitment in Ireland, as in most other European countries, which guides government strategy in the fields of education, health, transport and social welfare⁴. This is related to social justice considerations, whereby “everyone has the right to education and ... higher education shall be equally accessible to all on the basis of merit”⁵. It is based on the idea that a (more) meritocratic and cohesive society can be achieved by enabling young people to pursue educational credentials in line with their abilities, and the notion that this is not only right but also necessary in order to compete effectively in global markets. As Clancy (2015) observes, “[t]he unprecedented social demand for higher education is matched by a consensus among public policy makers that in our knowledge-rich society the higher education system is a key determinant of economic and social development”⁶. This is an important objective of the Bologna Process⁷, the set of European agreements currently driving change in higher education in Europe. A recent report by the HEA provides a similar synthesis:

The economic and social rationale for the continuing prioritisation of access remains strong and the HEA believes that the drivers for this policy are clear and unambiguous:

- *Widening access offers students the ability to improve their life chances in terms of citizenship and economic security*
- *A highly educated workforce supports the attraction and retention of high-end jobs*
- *It taps into the skills and talents of groups who have not traditionally progressed to higher education and has a positive ‘ripple’ effect within families and communities*
- *It fosters the ongoing development of a more engaged, inclusive and secure society*
- *It better positions Ireland to cope with the ongoing change, both economic and social, that is a feature of modern life.*⁸

These different objectives are not always perfectly aligned, as other commentators have observed⁹, and there are tensions between the ‘social justice’ arguments for equity in access

⁴ Clancy, P. (2015) *Irish Higher Education: A Comparative Perspective*. Dublin: IPA; Fleming, T., Loxley, A., Finnegan, F. (2017) *Access and Participation in Irish Higher Education*. London: Palgrave Macmillan.

⁵ UN General Assembly (1948) *Universal Declaration of Human Rights* (217 [III] A). Paris: UN, Article 26(1).

⁶ Clancy, P. (2015) *op. cit.*, p. 1.

⁷ European Union, Bologna Declaration, 1999 (see, for example, the 2007 London Communiqué); cf. Kooij, Y. (2015) *European Higher Education Policy and the Social Dimension: A Comparative Study of the Bologna Process*. London: Palgrave Macmillan.

⁸ HEA (2016) *Higher Education System Performance 2014–2016*. Dublin: Higher Education Authority, p. 27.

⁹ Clancy, P. (2015) *op. cit.*; Fleming et al. (2017) *op. cit.*

to higher education and the ‘human capital’ agenda, which emphasises the requirements of employers and the economic system. These tensions are found in many areas of policy-making in Ireland, due to the complex ways in which more traditional social Catholic and social democratic influences have been refracted in recent years through the increasingly dominant neoliberal prism. However, these different perspectives all emphasise the promotion of equity of access to higher education.

In recent decades, research in Ireland and other economically-advanced countries has demonstrated the existence of large and enduring disparities in educational participation and attainments according to socio-economic position. The survey carried out by O’Connell *et al.* for the HEA indicates that in 2004, 117 per cent of the ‘Professional Workers’ social class entered higher education, compared with just 38 per cent of the ‘Semi-skilled and Unskilled Manual’ social class and 36 per cent of the ‘Non-manual’ social class¹⁰. The more recent 2011 data presented in the National Plan 2015-19, which are based on different calculations and categories, suggest that 119 per cent of young people from the ‘Higher Professional’ socio-economic group, aged 18-20 years, enrol in higher education, compared with just 23 per cent for the ‘Non-manual Workers’ socio-economic group, 24 per cent for ‘Semi-skilled Manual’ and 25 per cent for the ‘Unskilled Manual’ group (the overall national participation rate is estimated at 52 per cent)¹¹. The fact that gender differentials in enrolment have been overcome in recent decades focuses even more attention on socio-economic inequalities and underlines the need to strengthen policies in this area and to intensify monitoring.

When monitoring equity of access to higher education, it is important to situate these disparities in relation to the education system as a whole. Social inequalities accumulate across the life cycle and develop incrementally within each level of the education system. It is necessary to assess participation in higher education by socio-economic position within relevant age cohorts, but this is not sufficient from a policy-making perspective. There may be large disparities in access to higher education, but these may be due, in part or entirely, to inequalities in the completion of second-level education or in relation to examination results. Without a comprehensive picture of the origins, distribution and nature of educational inequalities, it will not be possible to develop effective policies. This is an example of how effective policies require both an “evidence base” and “joined-up” or holistic forms of analysis¹².

‘Joined-up policies’ which seek to understand how inequalities are structured at a systemic level require a ‘joined-up’ evidence base. As far as higher education is concerned, this means analysing the interface between post-primary, further and higher education and seeking to understand how participation and achievement are structured in and across the different components of the education system. This expanded evidence base must identify the obstacles to enrolment, progression and transition in an integrated, coherent way, which

¹⁰ O’Connell, P., Clancy, D. and McCoy, S. (2006) Who Went to College in 2004? A National Survey of New Entrants to Higher Education. Dublin: HEA, p. 54.

¹¹ HEA (2015a) *op. cit.*, p. 43.

¹² See HEA (2015a), *op. cit.*, p. 16: “Equity of access policies must span the entire education spectrum and take a ‘whole of education’ approach to social inclusion”.

means following cohorts of students through the education system. It is not sufficient to rely on occasional surveys which provide a snapshot of the situation at a specific point in time, and which run the risk of providing a disjointed and contradictory picture. What is needed is an integrated and systemic approach to monitoring and studying educational inequalities on an ongoing, real-time basis¹³.

This kind of integrated system must satisfy a number of criteria. Firstly, it must follow students over time, so that we can make comparisons at key moments between those who leave the education system and those who remain, and those who choose certain pathways within it. Secondly, it must be possible to project forwards and backwards, in the first case to assess the impact of interventions like DEIS on participation in higher education, and in the second case to evaluate the early educational experience of those who enrol or choose not to enrol at university. Thirdly, it must contain reliable and comparable measures of key concepts and target groups. Fourthly, it must be efficient and cost-effective and fifthly, it must satisfy existing requirements in relation to data protection and security. Finally, it must be flexible enough to adapt to the evolving requirements of a dynamic field of public policy. These six criteria pose a serious challenge, and represent an ambitious “horizon” for the development of a new and more powerful monitoring framework.

It is helpful to think of the new data plan as contributing to the national data infrastructure, where this is viewed as a relatively stable, shared information resource which can serve several purposes and is of strategic policy importance. This enables us to appreciate how the proposed data plan relates to other aspects of the national data infrastructure, such as the system of PPSN codes, addresses and Eircodes, Ordnance Survey maps, Census Small Areas and the Census of Population itself. This infrastructure comprises a number of different elements, ranging from legislation to data protocols, public and private organisations, new and existing datasets and statistical definitions. This data plan is about bringing these different elements into a specific kind of configuration in order to provide policy-makers and administrators with reliable and accurate information on educational equality. The plan includes a realistic assessment of how to achieve this goal, by mapping out the “landscape” in relation to educational data, identifying potential ‘roadblocks’, on the one hand, and ‘enabling factors’, on the other.

When viewed from this perspective, the data plan necessarily touches on the activities, goals and interests of a number of different organisations, both public and private. Its implementation is a collective endeavour and requires the active participation of several stakeholders. This means that its key components, data requirements and other elements must be agreed with other interested parties and must satisfy their knowledge requirements for the foreseeable future. Given the broad nature of the data plan, and its wide-ranging coverage, it is appropriate that it should be supported and promoted at a high level, and not just by the HEA. The policy importance of equity of access should motivate a strategic commitment on the part of the Government to developing the national data infrastructure

¹³ In addition to social disparities in entry to higher education, there are significant differentials in relation to retention/progression, forms of higher education and choice of course. In equalities in relation to these processes are likely to become increasingly important in coming years, particularly if participation rates continue to increase.

in such a way as to facilitate the required ongoing monitoring activities in the sphere of education. The initial impetus for this development should thus come from Government, and should include the legislation and resources necessary to motivate and legitimate the participation of relevant bodies. In this sense, the data plan outlines a collective endeavour with shared objectives which is coordinated and encouraged from above by targeted government interventions. Recent developments in relation to data protection – discussed below – make it even more important and urgent to identify the required ‘enabling legislation’.

The core of the data plan is based on linking two key databases, namely the DES Post-Primary Online Database (which contains a comprehensive register of post-primary school students in Ireland) and the HEA Student Record System (which contains a comprehensive register of students at publicly-funded Higher Education Institutions in Ireland). It is proposed to link these datasets at the individual level using PPSN numbers, validated using address information and Eircodes. If this is not possible due to data protection or other considerations, these two datasets could be linked at the aggregate level. This may represent an appropriate intermediate step in the implementation of the data plan.

The family address of students in both datasets will have to be geocoded, with the aim of linking them to a Census Small Area and associating them with up-to-date census data at this fine level of spatial definition. This will most likely have to be carried out by each organisation separately¹⁴. Socio-economic position will be measured using a Small Area deprivation index, following the procedures pioneered by health researchers in the 1980s and 1990s, which are also used by the Department of Education to identify DEIS schools, and by the HEAR scheme since 2009 as part of a multi-indicator approach to assessing eligibility¹⁵. Additional information could, in time, be integrated from other sources such as the CAO (applications), UCAS (applications to HEIs in the UK), SUSI (student grant system) and SOLAS (relating to further education and training). The organisations concerned have developed powerful databases which provide information on various aspects of educational pathways and transitions which are of great relevance to the assessment of educational inequalities.

In the ‘optimal’, individual-level approach, the entire P-POD database would be matched with DES and SEC data to identify students who completed the Leaving Certificate and achieved specific results. P-POD contains information on country of birth, ethnicity and nationality, facilitating the identification of migration background and Travellers. The census would then provide the crucial proxy measure of socio-economic position (the Pobal HP Deprivation Index). This combined dataset could, in the future, be matched with individual-level CAO and UCAS data, to identify successful applicants to higher education institutions in Ireland, Northern Ireland and Great Britain. Students making a direct transition, deferring for 1 year, deferring for 2-4 years or deferring for 5 years or more could be identified on the

¹⁴ From this point of view, it is relevant to note that the HEA has decided to collect address and Eircode information on all students in the SRS, and higher education institutions will have responsibility for collecting these data and providing them to the HEA from March 2018.

¹⁵ Byrne, D., Doris, A., Sweetman, O. and Casey, R. (2013) *An Evaluation of the HEAR and DARE Supplementary Admission Routes to Higher Education*. Maynooth: NUI Maynooth.

basis of the year in which they left school or sat the Leaving Certificate, which could be provided by the DES. SRS data could be used to identify part-time and flexible learning students, those coming from the further education sector and the Fund for Students with Disabilities could be used to identify those with a disability.

At the core of this approach is the merging of data from individual-level datasets which contain personal identifiers. Each organisation has a different way of identifying individual students, and there is therefore a need to work towards the definition of a 'minimal' set of personal information that should be collected in a standardised way by each body, for identification purposes, as well as protocols for collecting these data and for validating PPS numbers and Eircodes. We believe that each body should collect and validate an appropriate set of personal information, as well as using their own ID numbers.

During this process, considerable attention should be paid to the collection of accurate and well-defined address information. This process appears to be beguilingly simple, but actually conceals considerable difficulties and risks. The first address collected should be the "Address of parents/guardians", whilst the second should be identified as "where the student currently resides (if different from above)". For mature students, it may seem rather incongruous to provide the parents' or guardians' address, as the individuals concerned may be economically independent and living away from home for many years. Nevertheless, collection of both addresses is warranted in this case as well, to facilitate the computation of transition rates that take account of social origins.

The direct exchange of datasets between organisations may raise insurmountable difficulties regarding data protection, security and access. The data plan thus foresees a role for a 'trusted third party', effectively centralising the process of data linking and anonymisation and regulating data access and analysis. The trusted third party would be responsible for encrypting the component datasets, linking databases, anonymising individual records, making custom datasets available for analysis in a secure data environment, checking all output to ensure that no sensitive information is disclosed and destroying all files at the appropriate point in time. This will help to reassure data providers that commitments to clients (in terms of data protection) and to society as a whole (in terms of monitoring social equity) are satisfied without undue risk. It would arguably be most appropriate for this role to be assumed by the Central Statistics Office, given its statutory role in relation to the "collection, compilation, extraction and dissemination for statistical purposes of information relating to economic, social and general activities and conditions in the State"¹⁶. Most of the stakeholders have already signed agreements with the CSO to provide individual-level data to the Administrative Data Centre.

The main reason for developing a plan based on linking and analysing information from administrative databases is to avoid the costs, interpretational difficulties and burden associated with existing data collection procedures¹⁷. Not only is it increasingly onerous to collect questionnaire data, it is also increasingly problematic due to selective non-response,

¹⁶ See the website <http://www.cso.ie/en/aboutus/descriptionsandfunctions/csomandateandmission/> [consulted 19 September 2017].

¹⁷ These are described in detail in the next section.

problems with occupational coding, social mobility, demographic transformations and associated difficulties of interpretation. Alternative approaches based on linked administrative datasets have been adopted with greater frequency in recent years in Ireland and the UK, precisely because they allow us to answer more questions, to gain greater insights and to obtain more accurate estimates¹⁸. Although they pose challenges, they are increasingly feasible in technical terms due to the development of powerful datasets containing information on service users across different sectors of policy delivery.

Existing data collection procedures for monitoring equity of access should be maintained whilst putting the key elements of the new educational data infrastructure in place. This will minimise risks whilst maximising opportunities for learning as we compare the results of new and old approaches. It is important to build on the positive aspects of the existing methodology – which has served policy-makers well and has contributed to the reduction in socio-economic inequalities in recent years, which Clancy (2015) documents in his recent book.

In the following sections, we will address the issues raised by the new data plan in greater detail, with a view to demonstrating its feasibility and potential. We will address the challenges regarding data protection and provide a rich discussion of substantive issues like the status of mature students, international flows of students and the measurement of socio-economic position by Small Area proxy. The discussion is confined to the key issues, with a view to providing readers with concise information on topics that are of particular interest to them. We begin, however, by providing some background information on the evolution of policies for equal access to HE and the specificities of the Irish system.

¹⁸ Particularly relevant, in this respect, is the recent institution of an “Administrative Data Research Network”, based at the University of Essex, which carries out data linkage and anonymisation for the research community. Amongst other datasets, the ADRN has negotiated access to the national pupil dataset, applications to higher education (UCAS) and the student record system, which is held by the UK Higher Education Statistics Agency. Rather than holding these datasets, the ADRN negotiates access with the relevant Government agency and acts as ‘trusted third party’, ensuring that data security requirements are respected.

5. Background

As noted in the previous section, the need to pursue equity of access is grounded in the individual and collective benefits that derive from this, which has become a common theme in education policy in OECD countries over the past two decades:

As well as the benefits that individuals derive from higher education, there are also wider economic and social benefits – our educated workforce is Ireland’s greatest economic asset, and we need more people to take up higher education to drive economic progress. We are now beginning to see skills shortages in vital and dynamic areas of our economy such as ICT and pharma, and there is some evidence that other sectors too are beginning to experience skills shortages.¹⁹

Clancy (2015) outlines how successive European governments have come to recognise the potential of education to contribute to the achievement of social justice and how equality of access has become a key policy objective in most of the economically advanced countries. As part of the Higher Education Act 1971, the Higher Education Authority (HEA) is obliged to promote “the attainment of equality of opportunity in higher education” and under the Universities Act (1997), all institutions must “facilitate lifelong learning through the provision of adult and continuing education” and “promote gender balance and equality of opportunity among students and employees of the university”. Similarly, the Institutes of Technology Act (2006) identifies equity of access to higher education as a key objective, particularly by “economically or socially disadvantaged persons, by persons who have a disability and by persons from sections of society significantly under-represented in the student body”. Clancy provides an overview of the expansion of higher education in Ireland:

Full-time enrolments in the dominant state-aided sector increased almost seven-fold over the past four decades. The rate of increase was very consistent over the 1970s, 1980s and 1990s, with growth rates of 64%, 71% and 82% during these decades. In the 1990s the growth rate was 34%, although from a much higher base. These higher enrolments are reflected in constant increases in the rate of admission. In the late 1960s the rate of admission was 5%. The pattern of entry to higher education has been monitored over a period of more than two decades in a series of reports commissioned by the Higher Education Authority (HEA) (Clancy, 1982, 1988, 1995, 2001; O’Connell et al., 2006). These reports reveal admission rates into full-time higher education of 20% in 1980, rising to 55% in 2004, and subsequently to about 65%.²⁰

The National Office of Equity of Access to Higher Education was established in 2003 in the HEA and co-ordinated work on the development and implementation of National Access Plans for the periods 2005-7 and 2008-13. The current plan – the National Plan for Equity of Access to Higher Education (2015-19) – was published following wide-ranging consultation with stakeholders and aims to “ensure that the student body entering, participating in and

¹⁹ Higher Education Authority (2015a) *op. cit.*, p. 15.

²⁰ Clancy, P. (2015) *op. cit.*

completing higher education at all levels reflects the diversity and social mix of Ireland's population".

For the duration of the National Access Plan, the DES and the HEA are committed to increasing participation in higher education by the following under-represented groups²¹:

- Entrants from socio-economic groups that have low participation in higher education
- Students with disabilities
- First time, mature student entrants
- Part-time/flexible learners
- Further education award holders
- Irish Travellers

The definition of these target groups is intended to enable the HEA and the DES to assess progress and to assist institutions in developing practical actions to promote access. The target groups form the basis of agreements with higher education institutions as part of annual performance assessments by the HEA. The list of under-represented groups should thus be seen as an evolving component of the system, which means that the data plan should be able to accommodate new groups as they are added (e.g. young people from lone parent families or members of specific ethnic minorities²²). One study estimates that 17 per cent of undergraduates are parents, with a much larger share of student parents being observed among mature, part-time and flexible learning students²³.

The HEA is currently developing an overall data development and knowledge management strategy for the higher education sector, and the data collection and analysis initiatives undertaken as part of the National Access Plan form part of that strategy. Over the duration of the plan, the HEA aims to further develop data collection and analysis methodologies so

²¹ In 2015/16, there were 43,460 new full-time entrants to higher education. Of the 222,618 higher education students, 80% were full-time, 17% part-time and 3% distance-learning students. 5% of new entrants were aged 30 and over, and 13% were aged 23 and over. 7.8% of new entrants had a disability (57.4% a learning difficulty, 29.4% a psychological or emotional condition, 21.8% a chronic condition, 8% a physical disability and 5.8% a sensory disability). Just 0.1% of respondents to the Equal Access Survey were Irish Travellers, 1.1% were Africans, 0.9% Chinese and 1.8% Other Asian. For further background information on the composition of higher education, see HEA (2017a) *Key Facts and Figures. Higher Education 2015-16*. Dublin: Higher Education Authority.

²² Clancy (2015) notes that minority ethnic group students are still concentrated in primary and secondary schools, whose share of immigrant students almost trebled in 10 years to reach 10-12% by 2006. As far as lone parents are concerned, Byrne and Murray (2017) note that "Target setting has been hampered, as baseline data on participation by lone parents in 2008 and 2015 was unavailable and therefore a target could not be set. However, the National Plan for Equity of Access 2015-2019 contains specific actions for improving the systematic collection of relevant, comparable data necessary for improving the evidence base for access policy. A Data Plan for Equity of Access to Higher Education is currently being developed, and lone parents should be a key consideration in this process"; Byrne, D. and Murray, C. (2017) *An Independent Review to Identify the Supports and Barriers for Lone Parents in Accessing Higher Education and to Examine Measures to Increase Participation*. Maynooth: Maynooth University, p. 10.

²³ Byrne and Murray (2017) *op. cit.*

that these can underpin policies and deepen knowledge of target groups. In relation to data collection, the National Plan aims “to gather accurate data and evidence on access and participation and to base policy on what that data tells us” (Goal 3). This policy is in line with this data plan, and is likely to involve both harmonisation of definitions and agreement on a fundamental set of indicators to be used for monitoring purposes. As administrative data sources are used for monitoring purposes, a communication strategy will also be required to ensure that citizens are informed about the ways in which the data they provide will be used. If a specific dataset is to be used to monitor equity of access, it will be necessary to inform those who contribute information to that dataset that their data will be used for this purpose.

The HEA has gathered data on entrants into higher level education since 1982. For the past three decades, the main approach used to assess equity of access for this group has drawn on occupational groups and the methodology to measure under-representation by particular socio-economic groups developed by Professor Patrick Clancy. Professor Clancy’s studies were based on data from Student Record forms, from the CAO and from a voluntary survey issued to all students offered a place via the CAO/CAS²⁴. Data were also obtained from the Central Statistics Office and the Department of Education to provide a context for the survey, and there were some variations over time in data collection procedures. As these studies were produced, they became an integral part of policy development for equity of access in higher education and shed light on the nature of persistent inequalities and identified improvements in participation as they occurred²⁵.

The socio-economic background of students was measured using 1) parents’ principal economic status 2) parents’ socio-economic group and 3) parents’ social class. Social class position and socio-economic group were determined by coding descriptions of occupations for both parents, in line with the “dominance approach” which assigns the whole family to the social class category of the parent or guardian with the highest position²⁶. In 2007, all publicly-funded higher education institutions started to gather information by means of a voluntary survey coordinated by the HEA on the socio-economic, ethnic/cultural and

²⁴ Clancy, P. (1982) *Participation in Higher Education*. Dublin: Higher Education Authority; Clancy, P. (1988) *Who Goes to College?* Dublin: Higher Education Authority; Clancy, P. (1995) *Access to College: Patterns of Continuity and Change*. Dublin: Higher Education Authority; Clancy, P. (2001) *College Entry in Focus: A Fourth National Survey of Access to Higher Education*. Dublin: Higher Education Authority; see also O’Connell *et al.* (2006) *op. cit.*; Clancy, P., Hayden, J., Kelleher, M. (2010) *External Audit of Equal Access Survey*. Dublin: Higher Education Authority.

²⁵ “Optimists will point to the more than three-fold increase of the lower socio-economic groups from 8% to 29% [between 1980 and 1998], while pessimists will note that the percentage-point differences between the two aggregate groups have increased over the 18-year period, from 27 (35%-8%) to 30 (59%-29%). It is suggested that the best way to measure changing inequalities is to calculate odds ratios (Clancy, 2001: 178-9). These calculations suggest that there has been a reduction in socio-economic group inequalities of about 39% over the 18-year period.” (Clancy, 2015, *op. cit.*).

²⁶ Students indicate the occupation of their parents or guardians as part of the survey, with reference to their current or most recent job. Data are coded externally and aggregated to 11 socio-economic groups (including unknown). The target socio-economic groups for the HEA are currently ‘Non-manual Workers’ and ‘Semi-skilled and Unskilled Manual Workers (including agricultural workers)’.

disability background of new entrants. Responses to this “Equal Access Survey”, which facilitates ongoing monitoring of equity of access, are provided to the HEA as part of the annual data return by HEIs for the Student Records System (SRS)²⁷.

The Equal Access Survey is repeated each year and uses an adapted version of Clancy’s methodology. It seeks to measure participation rates for full-time, undergraduate new entrants to higher education, aged 18 to 20 years, from relevant socioeconomic groups, as a proportion of young people from the same age cohort and the same socioeconomic groups in the national population (as measured by the Census of Population, taking into account the interval between the last Census and the current Equal Access Survey)²⁸.

In 2010, an external audit of the EAS was carried out to assess the effectiveness of the Equal Access Survey as a means of assessing equity of access to higher education and allocating funding to support Access initiatives²⁹. The audit concluded that the data provided by the survey were sufficiently robust to satisfy this purpose, but had a number of shortcomings, including the voluntary character of the survey and consequent non-response. Clancy observes that despite intensive data collection efforts since the 1980s, “[w]hile good comparative data are available on the elimination of quantitative inequalities in the access of women to higher education and also on the extent of (persisting) generational inequalities, we remain very poorly informed on the changes in social group inequalities and on changing inequalities by ethnic groups and by disability.”³⁰.

The Equal Access Survey was discussed during the consultation phase for the National Access Plan 2015-19. Among the points made was the need for a more refined, multi-indicator approach to how different areas of disadvantage are targeted. In addition, with advancements in technology, it was recognised that data collection and data analytics have gained in scope and that important data sources, maintained by the HEA and other organisations, could have the potential to shed light on access considerations. Whilst in the recent past, *ad hoc* surveys were the only way to obtain data on inequalities in access to higher education, this situation has now changed.

²⁷ Core funding for higher education institutions is now based, in part, on access considerations. The Recurrent Grant Allocation Model (RGAM) is based primarily on student numbers, but this is weighted by type of course, full-time or part-time status, undergraduate or postgraduate, taught or research students and the representation of specific target groups. For example, under-represented socio-economic groups, first-time mature students and Irish Travellers are funded entirely for their first two years, and students with disabilities are funded for their entire course duration, with a further multiplier of two applied to recognise support costs. Core funding support for improving access to higher education involves an additional weighting premium of 0.33 for the discipline-based weighting for all eligible access students.

²⁸ By contrast, Clancy calculated participation rates using all age cohorts of students (including mature students) and expressing them as a percentage of the age cohort to which the majority of new entrants belonged.

²⁹ See Clancy et al. (2010) *op. cit.* Over €47m was allocated in 2017 to HEIs on the basis of “Equal Access” considerations. Data from the Equal Access Survey are also used to inform the allocation of the Student Assistance Fund, which amounted to €8m in 2016/17, and 50 per cent of the allocation to HEIs was based on the number of students from under-represented socio-economic groups.

³⁰ Clancy (2015) *op. cit.*, p. 66.

A crucial enabling factor is the way in which the national data infrastructure has been enriched in recent years as a result of the following developments:

- (i) The creation of census Small Areas for the 2011 census, and the consequent availability of the Pobal HP Deprivation Index at this level
- (ii) The GeoDirectory, which provides an up-to-date register of the XY coordinates of addresses and thus facilitates geocoding to Small Area
- (iii) The introduction of Eircodes (in 2015), which uniquely define each residential and business address
- (iv) The implementation of new administrative databases for recording information on primary and post-primary school students (in 2014)
- (v) The maintenance of a quinquennial census ensures that spatial data are available on a more timely basis than in most other European countries
- (vi) The implementation of computerised datasets by the HEA, DES, CAO, CDETb and SOLAS, containing up-to-date information on third-level applications, incentives for overcoming the effects of disadvantage and disability on educational participation, student grants and both further and higher education student records

In the next section, we will describe the ‘logic model’ which is at the heart of this data plan and show how this can guide the construction of a new system for monitoring equity of access to higher education.

6. Logic Model

When assessing access to higher education, it is helpful to distinguish between two different research designs. The first focuses on the characteristics of students entering higher education, and compares them with a reference group. In line with this approach, it is possible to calculate the percentage of students from less affluent families enrolling in higher education in relation to their share of the population. This is the approach adopted by Professor Clancy in his access studies and by the HEA from 2007.

The second type of analysis involves studying the relationship between students enrolled at HEIs and the actual population from which they are drawn. In this type of analysis, educational outcomes are treated as the result of a series of selections from an initial population. It is possible to define these selections in such a way that they are nested and incremental, forming a chain in which each selection operates on the basis of the previous one. For example, it is possible to view students who apply to higher education after post-primary education as a subset of those who completed the post-primary cycle, where the latter represent a subset of those who attended school to age 16 and so on. These selections are probabilistic and influenced by a range of different factors and characteristics related to the individual, family, neighbourhood, school and peer group.

This provides a flexible model of educational trajectories that can help to orient the data plan and clarify the measures and information required. It shows how the differentials observed at any point in the chain are dependent on those generated at earlier stages. It also brings the complexities into focus: precocious exit and later re-entry to the education system can be viewed as qualifying or compensating for initial inequalities. In this way, overall equity may be viewed as a function of ‘direct’ and ‘delayed’ transitions. As we mentioned earlier, survey data on participants in higher education can only provide summary information on the extent of inequalities, but cannot shed light on the underlying mechanisms involved. In order to develop effective policies, it is necessary to acquire and analyse data relating to a range of different transitions. It would be costly, onerous and difficult to conduct this kind of analysis each year using questionnaire surveys. Fortunately, the development of powerful and comprehensive administrative databases provides an effective way of calculating these transitions.

This means that the research agenda in relation to equity of access to higher education increasingly overlaps with the research agenda regarding educational outcomes more generally, and this is why we referred in the previous section to a “shared research agenda” and the need to develop the national data infrastructure to strengthen analytical capacity. By using administrative databases we can identify the actual populations from which each selection is made, which eliminates one of the largest sources of error inherent in the survey-based approach. This shift in emphasis creates economies of scale and opens up many new possibilities for understanding educational transitions. It reduces burden on higher education institutions and avoids the expense of collecting new data.

It is important to situate the study of equity of access to higher education within a broader model of access, participation and achievement that looks at how the education system as a whole performs from a social inclusion perspective. This provides the rationale for pursuing a joint data plan and a joint analytical strategy, as all stakeholders may find that they can only answer their specific research and policy-related questions by coordinating data

collection and analysis with other organisations. The growing emphasis on joined-up and inter-sectoral approaches to policy-making must therefore be combined with a joined-up approach to data and research. The incremental and phased introduction of this data plan will enable policy-makers and stakeholders to gain an understanding of how it can help them to fulfil their statutory objectives, avoiding difficulties, gaps or uncertainties in the passage from one monitoring system to another.

The new assessment framework is based on the theoretical model of educational transitions that is summarised in the Logic Model below. This provides a graphical representation of the relationship between the different elements that influence access to higher education and constitutes a 'road map' for building consensus and implementing the new data plan. The Logic Model shows the key educational transitions that are involved in differential access to higher education. In the future, this specific focus can be complemented by more 'distal' measures of equity relating to post-primary education ('upstream equity issues') and subsequent access to postgraduate study and the labour market ('downstream equity issues')³¹. The Logic Model provides a graphical 'roadmap' which carries out several functions, including (a) helping to build consensus amongst stakeholders in relation to the aims and characteristics of the data plan, (b) assisting in the conceptualisation of different forms of inequality in access to higher education, (c) identifying necessary sources of data.

The resulting trajectories are complex and comprise discontinuities and interruptions, as well as orderly progressions from one level to the next. The probability of completing each transition is influenced not only by the characteristics of the individual, their family and the neighbourhood, but also by their previous transitions and experiences, giving rise to an incremental process. For example, access to higher education is dependent on applying to a HEI and on achieving the required points, which is dependent on Leaving Cert results, which are dependent on school experience and so on. This means that disadvantage accumulates over time and inequalities are distributed across the education system.

It is also worth noting that inequalities in relation to one transition may be explained by previous transitions, as when the under-representation of a particular social group at university is due to a tendency to leave school before completing the Leaving Certificate³². It is also important to stress the difference between inequalities that are due to explicit selection mechanisms and inequalities that are due to choices or informal selections. For example, the under-representation of people from disadvantaged areas, within the university system, may be due to the difficulties they face in meeting entrance requirements (selection) or could be due to a tendency to seek work after completing school (preference).

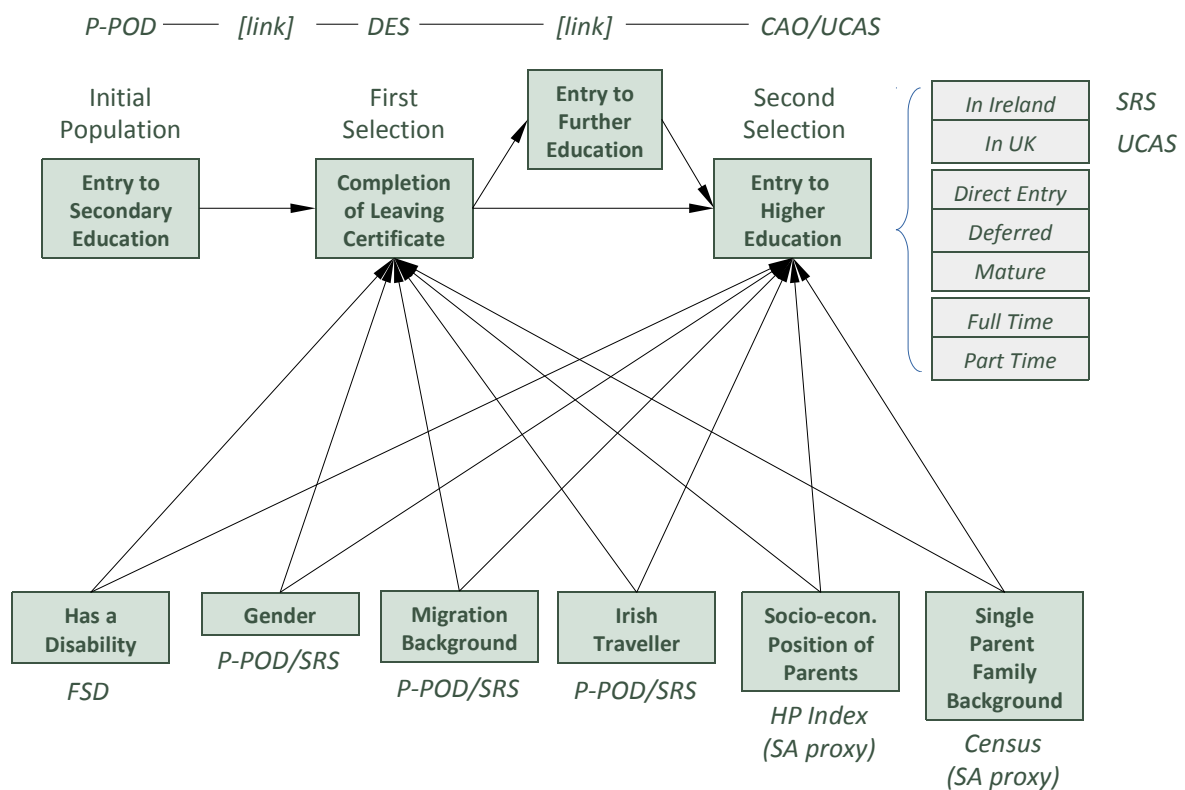
³¹ It is worth noting, from this perspective, that the HEA is already involved with the CSO in studying graduate outcomes by using linked data. See HEA (2017b) *What do graduates do? The class of 2015*. Dublin: HEA.

³² It should be noted, of course, that a relatively large proportion (in the region of 90%) of Irish pupils complete post-primary education, and of these a similar proportion sit a Leaving Certificate exam that (at least in principle) provides access to higher education.

The logic model shows that the probability of entering higher education should be assessed in various ways and for a range of social groups:

- (i) as part of a chain of transitions: in relation to post-primary students, in relation to those who completed the Leaving Certificate and in relation to those who attained a minimal set of grades at the Leaving Certificate
- (ii) for Irish institutions alone, and then for all higher education institutions in Ireland, Northern Ireland and Great Britain
- (iii) for students making a direct transition from school to higher education, those who defer entry for 2-4 years and those who enter as mature students
- (iv) for full-time versus part-time students
- (v) in relation to disability, gender, migration background, ethnic identity, socio-economic position and family composition
- (vi) meaningful interactions between these factors should also be explored, with a view to determining whether inequalities in relation to full-time/part-time status are explained by socio-economic group, for example

Figure 1 **Logic Model**



Our key aim may now be reformulated in light of the Logic Model: at key points along the educational pathways illustrated in the model, we can estimate the probability of making transitions, comparing across the main target groups defined in the National Plan as well as students from lone-parent families, women and ethnic minorities.

7. Assessment of existing approach to measuring equity of access

In this section we will describe current procedures for monitoring equity of access and explore their shortcomings. It is important to clarify these issues in order to define a data plan which can yield better estimates of socio-economic and socio-demographic differentials in enrolment. Since 2007, equity of access to higher education has been assessed in the following way³³:

- (i) All first-time, full-time undergraduate entrants to higher education aged 18-20 years are asked to complete a voluntary questionnaire as part of the Equal Access Survey, and the data are provided to the HEA by publicly-funded HEIs as part of the Student Records System (SRS)
- (ii) A textual description of the parents' or guardians' occupational position is coded externally following CSO guidelines to obtain a classification by social class and by socio-economic group and family members are assigned to a group on the basis of the father's occupation, but also using the so-called "dominance" approach, which classifies family members by reference to the person with the highest occupational position³⁴
- (iii) The data are re-weighted using information provided by the CAO to reduce the bias generated by selective non-response
- (iv) The Census of Population is used to calculate the number of young people aged 18-20 years in each socio-economic group, after taking account of the number of years since these census data were collected
- (v) The distribution of HE entrants between socio-economic groups is compared to the distribution of the general population, leading to an estimate of the proportion of young people from each SEG entering higher education

This process, comprising various kinds of data collection, coding, re-weighting and analysis, has both merits and drawbacks. In terms of merits, all operations are transparent and relatively easy to understand. When it was originally designed, it was arguably the only feasible and cost-effective way of assessing equity of access. The relevance of the comparison between the distribution of HE entrants and young people in the population is self-evident and directly linked to the National Plan. The socio-economic groups have a substantive interpretation and may be linked to existing sociological theories regarding the role of socio-economic position and social class in determining life chances. In addition to the census, other Irish datasets and surveys are quite often classified by socio-economic group, facilitating comparisons. Whenever new data are required, coding operations can be repeated in a predictable manner to yield comparable classifications for comparisons between different time-points. Some (limited) comparisons can be made with other countries using the percentage of people in the manual social classes, for example (see Clancy, 2015).

³³ As we noted earlier, there are strong similarities between the approach pioneered by Professor Clancy and that adopted by the HEA from 2007 (see O'Connell et al., 2006, *op. cit.*), and the difficulties identified here apply to both methods.

³⁴ Erikson, R. (1984) 'Social Class of Men, Women and Families', *Sociology*, 18, 4: 500-514.

There are, however, a number of difficulties:

- (i) It has proved increasingly difficult to gather information about socio-economic group and response rates are rather low³⁵
- (ii) Refusal to respond to questions about socio-economic group is not distributed in a random fashion and is likely to be influenced by socio-economic group itself, and it may not be sufficient to simply reweight the data
- (iii) The need to express data on entrants as a proportion of a population that was quantified in a census held between one and five years earlier reduces precision
- (iv) As the data used to calculate the numerator and denominator are gathered using two different processes (and are based on information provided by different people) there is a potential for bias³⁶
- (v) The need to define the denominator in terms of age alone gives rise to further problems, as it does not measure the population that *could potentially enter higher education* in any given year, which also means that it cannot pinpoint the source of inequalities
- (vi) The manual coding of individual socio-economic group is costly
- (vii) The coding of individual socio-economic group is not possible for a growing percentage of people and occupations, and the reasons for this are themselves related to socio-economic group
- (viii) Definitions of social class categories and socio-economic groups are subject to periodic revision, and are typically not comparable across jurisdictions
- (ix) The socio-economic group classification is not unidimensional, but comprises several underlying dimensions, making it difficult to interpret change
- (x) Processes of demographic change within the population can give rise to difficulties in interpreting change over time
- (xi) Processes of economic and occupational restructuring also give rise to difficulties in interpretation, as the size and composition of socio-economic groups can vary rapidly

The combined effects of these aspects of the current approach to monitoring equity of access make it difficult to provide a correct and unequivocal interpretation of change over time, particularly as a result of points (ix), (x) and (xi). For example, one of the findings of the ESRI report by O’Connell et al. (2006) is that the rate of participation in higher education for

³⁵ As low as 14.5% for Trinity College Dublin, with a national average of 57.5% in 2008-9 (see Clancy et al. (2010) *op. cit.*, p. 9).

³⁶ There are significant statistical anomalies, in this respect, as certain socio-economic groups consistently have estimated participation rates that are over 100%. For example, in 2011, the ‘Higher Professional’ socio-economic group had a participation rate of 119% (HEA, 2015a, *op. cit.*, Table A2). This suggests that not only is there potential for error when comparing data gathered in different ways at different points in time and from different individuals, but that there is also considerable bias. If the errors were evenly distributed, we would not expect consistently inflated participation rates for specific groups, such as these. It is therefore possible that children’s descriptions of their parents’ occupations differ systematically from their parents’ descriptions. If children’s descriptions tended to lead to the classification of their parents in higher socio-economic groups, then we could indeed observe participation rates over 100%. This risk may not have been given sufficient attention in official reports and reviews.

the 'Non-manual' socio-economic group declined between 1998 and 2004, from about 0.29 to about 0.26, making this the most disadvantaged category (see p. 9). At the same time, the participation rate for the 'Skilled Manual' group increased from 0.32 to roughly 0.55, and that for the 'Semi- and Unskilled Manual' category increased from 0.23 to about 0.36. The conclusion drawn by O'Connell *et al.* is that there have been considerable improvements in access for manual workers and own-account workers, whilst the 'Non-Manual' group has a lower (and falling) rate of participation³⁷. This group is quite large and heterogeneous, containing just under one-fifth of the population, and it includes a highly-disadvantaged subgroup within it. On the basis of this body of work, the 'Non-manual' socio-economic group was designated an under-represented category and incentives were put in place to support the participation of members of this group³⁸.

These conclusions are based on the assumption that changes in participation rates at different points in time relate to socio-economic groups whose characteristics are relatively stable. This assumption is problematic because the period between 1998 and 2004 was characterised by two other major processes of change which influenced the composition of the categories themselves, and hence their participation rates. The first of these was the increase in the proportion of lone parent households. In 1996, there were 49,819 lone mothers with children under 15 years of age, according to the Census of Population. This figure had risen to 90,818 by 2006, which represents roughly one-fifth of all households with young children³⁹. As the 'Non-Manual' socio-economic group comprises occupations which are sex-typed as 'female' (such as secretaries, hairdressers, waitresses etc.), the deterioration observed in the participation rate for this socio-economic group is quite possibly due to the fact that it increasingly coincides with the category of female-headed families (and lone mothers, in particular)⁴⁰.

There are considerable differences between two-parent and female-headed lone parent families in relation to social background and access to financial, social and cultural resources, and these differences could explain the low rate of participation that is observed for this socio-economic group. For example, the pressure to leave school and to look for paid employment as soon as possible is likely to be particularly strong for young people from

³⁷ As noted earlier, the data presented as an appendix to the National Plan (HEA, 2015a) provide a similar picture for 2012, although these are based on the HEA's revised approach and are therefore not directly comparable.

³⁸ Explicit targets were set for the 'Non-manual' socio-economic group in HEA (2010) *National Plan to Achieve Equity of Access to Higher Education 2008-2013: Mid-term Review*. Dublin: Higher Education Authority.

³⁹ For 1996, we used Tables A0331 and A0332, whilst for 2006, we used Tables C0331 and C0332. In relation to participation by lone parents themselves, see Byrne, D. and Murray, C. (2017) *op. cit.*

⁴⁰ The 'Non-Manual' socio-economic group comprised 613,285 individuals in 1996 and 818,573 ten years later (out of a total of 3,626,087 in 1996 and 4,239,848 in 2006). There were 794,521 households in Ireland in 1996 (Census Table A0333) and 1,006,850 in 2006, which implies that there were 4.56 individuals per household in 1996 and 5.18 in 2006. On average, therefore, there were at least 134,492 families in the 'Non-Manual' group in 1996, and at least 158,025 in 2006. It is quite possible, therefore, that lone mothers accounted for a significantly larger share of this category in 2006, although further analyses using individual-level data would be needed in order to confirm this hypothesis.

disadvantaged lone parent families. As Byrne and Murray (2017, *op. cit.*, p. 11) observe, lone parents are consistently identified in the literature as a group with relatively low levels of higher education and their families are generally exposed to a high risk of social and financial exclusion and vulnerability. Two-parent families can potentially have two sources of income, and can draw on more resources when organising child care (both parents can contribute, both families can assist). Rather than targeting *all* young people from the ‘Non-Manual’ socio-economic group, it could be argued that policies should focus more specifically on encouraging children from lone parent families to complete post-primary education, to obtain good grades and to apply for a place at a higher education institution⁴¹. These different interpretations of the same data lead to quite different policy conclusions, depending on whether we identify casualised service workers or lone mothers as the key target group.

This specific issue was addressed in an ESRI report, although it did not consider the specific factors discussed above⁴². The authors conclude that the Non-Manual socio-economic group has slipped below manual workers in terms of socio-economic position, interrupting the traditional ordering of occupations:

What has emerged is a previously unknown sub-group, representing nearly 10 per cent of the population, for whom a complex interplay of social, cultural and economic processes have led to low levels of participation in HE. The picture is clearly one of hidden disadvantage, as the merging of intermediate and other non-manual groups in most previous empirical work has concealed a dramatic picture of educational disadvantage among young people from other non-manual backgrounds. (McCoy et al., 2010, p. 167)

Another process was also taking place at the same time, namely the economic boom of the ‘Celtic Tiger’. This led to high rates of social mobility, as the demand for labour enabled workers to move from relatively less skilled to relatively more skilled jobs. Between 1996 and 2002, the number of manual workers (excluding agriculture) with children aged 15-17 declined by more than one third, whilst agricultural workers declined by more than half (O’Connell et al., 2006, *op. cit.*, p. 45). Over the same period, the number of people classified in the ‘Lower Professional’ socio-economic group increased by 21.3 per cent⁴³. In other words, many of the families of young people who enrolled in higher education in 2004 had already benefited from the opportunities generated by the economic boom.

As a result of this process, many of the families assigned to the professional and managerial classes in 2004 were socially mobile and included a reference person who had recently

⁴¹ There are therefore two distinct target groups that may deserve further attention in the future, from the perspective of the risk of educational disadvantage: lone parents themselves and their children. This represents a measurement challenge, as the exact composition of students’ families is not currently recorded systematically in any of the existing databases.

⁴² McCoy, S., Byrne, D., O’Connell, P., Kelly, E., Doherty, C. (2010) Hidden Disadvantage? A Study on the Low Participation in Higher Education by the Non-Manual Group. Dublin: HEA.

⁴³ These changes are merely reported by O’Connell et al. (2006), *op. cit.* and are not referred to when interpreting changes in participation rates.

moved out of the manual classes. This had the effect of lowering participation rates amongst the higher socio-economic groups, as their new members did not possess the accumulated social and cultural advantages that longer-standing members of the professional and managerial socio-economic groups typically possess. In other words, a considerable proportion of the higher socio-economic group categories in 2004 were drawn from groups which historically had lower rates of participation in higher education. What we find, in fact, is that participation rates declined between 1998 and 2004 for all non-manual socio-economic groups. These changes involve an irremediable confounding of different processes, making it very difficult to draw reliable inferences regarding the impact of policies and the need for interventions. In conclusion, we cannot assume that the characteristics of these socio-economic groups remain stable over time, which means that we cannot interpret changes in participation rates as expressing intrinsic attributes of these groups.

Another difficulty relates to the growing share of the population which cannot be assigned to any socio-economic group on the basis of information about the current or last job carried out by the household reference person. This is a well-known phenomenon in Britain and Ireland, and includes households where the reference person has never held a stable job, has a long-term limiting illness that prevents them from working, is institutionalised, not present, deceased or has an occupational status that is not clearly known to the respondent⁴⁴. This means that many of the most disadvantaged families cannot be assigned a socio-economic group and their low rates of participation in higher education are underestimated, as they are either excluded from analyses of access or combined with other groups (such as those performing newer occupations that are difficult to code).

The difficulties documented above risk undermining confidence in the evidence base for equal access policies, leading to a risk of adopting inappropriate policies or even abandoning the principle of evidence-based policy-making⁴⁵. Any new data plan for monitoring equity of access must therefore seek to overcome or avoid these difficulties.

⁴⁴ Forsyth, A., Furlong, A. (2000) *Socioeconomic Disadvantage and Access to Higher Education*. Bristol: Policy Press.

⁴⁵ The National Plan observes that “Changing perceptions and definitions of different socio-economic groupings, however, do present challenges in how to interpret such data. For example, there has been an increase in the numbers of students whose socio-economic group is ‘unknown’, and there are also difficulties of definition in relation to the profiling of the non-manual group in terms of disadvantage. The HEA is working to address these challenges and has implemented some solutions” (HEA, 2015a, *op. cit.*, p. 20).

8. Small-area deprivation indices as a proxy for social class

A key issue for the new data plan relates to the measurement of socio-economic position. In the absence of questionnaire data, how can this be measured? As we argued above, the Equal Access Survey does not yield a very reliable measure, due to selective non-response, the impossibility of coding up to one-fifth of respondents, due to the difficulties involved in combining the responses of fathers and mothers, comparing them with children's responses and accounting for family type and economic restructuring. Moreover, international comparisons are limited using the current methodology. This implies that there may be scope for improving on this measure.

One possibility is to use a proxy measure of socio-economic position by geocoding the residential addresses of students' families. By linking each student to a Small Area of residence, it would be possible to use census data as a proxy for socio-economic position. This would yield an indicator of relative position, and could be based on multiple indicators covering educational attainments, labour market situation, material affluence and disadvantage and demographic processes. The Pobal HP Small Area Deprivation Index provides such a summary measure and is already widely used in other areas of policy-making. This index is now used by the DES when geocoding the residential addresses of school students and this procedure will continue to be applied in the future, yielding an up-to-date geocoded database. It has also been used for a number of years as a criterion for obtaining support under the Higher Education Access Route (HEAR) scheme, which promotes access to higher education for students from disadvantaged backgrounds. This way of measuring socio-economic position yields a number of benefits, including:

- (i) Little or no loss of data and very low related bias
- (ii) An up-to-date denominator for calculating participation rates in higher education
- (iii) Reduced burden on HEIs in terms of data collection and coding
- (iv) Information on the socio-economic position of all post-primary students, removing the need to estimate this independently and inaccurately from the census
- (v) Possibility of following students through transitions within post-primary education (including continuation to Leaving Certification, examination results etc.) and after leaving school

If we can obtain deprivation scores for new entrants to higher education and (potential) applicants making the transition from post-primary education, we can calculate participation rates by deciles (ten equally-sized categories) or quintiles (five equally-sized categories) of deprivation scores. Although these categories are essentially arbitrary, they capture precisely what is needed in relation to equity of access by providing a measure of relative socio-economic position. One of the weaknesses of occupational categories – the lack of scale or hierarchy – is immediately overcome. The highest and lowest categories do not require interpretation, and any differentials that exist between them can be interpreted in a straightforward way as reflecting social inequalities. Targeted policy interventions would also be facilitated, as applicants from disadvantaged Small Areas could be encouraged and

supported, just as higher education institutions could receive an incentive for enrolling them⁴⁶.

The use of deprivation scores as a proxy for socio-economic position via the geocoding of addresses is well-established in epidemiology and health research (amongst other fields), where it has been shown to provide reliable results, particularly where Small Areas are used (rather than larger and more mixed geographical areas)⁴⁷. The Pobal HP Index appears to be well-suited to this task and much more appropriate than the ward-level indices (IMD, Polar3) which are used in the UK for similar purposes⁴⁸. When analysing individual survey-level data on health, well-being, child development and education social gradients are often assessed in this way, and this has proved to be particularly useful when seeking to measure inequalities⁴⁹.

When using aggregate data as a proxy for individual-level characteristics, heterogeneity is sometimes cited as a crucial problem. For example, college applicants from deprived areas could (in theory) come from affluent families⁵⁰. This is a possibility, and one which increases with the size and heterogeneity of the spatial units. The use of Small Areas considerably reduces this risk, as they were designed to be internally homogeneous and their small size enables them to achieve this goal. It is also worth remembering, however, that proxy data are used in this context to *estimate* equity of access by aggregating across large numbers of individuals. This means that misclassification would have to be highly asymmetrical before it could significantly influence the results; errors tend to cancel out between disadvantaged and affluent areas⁵¹.

⁴⁶ This is already partly the case, as one of the criteria for qualifying for the HEAR scheme is to be from a disadvantaged area, according to the HP Deprivation Index.

⁴⁷ See Domínguez-Berjón, F. (2005) 'The Usefulness of Area-Based Socioeconomic Measures to Monitor Social Inequalities in Health in Southern Europe', *European Journal of Public Health*, 16(1): 54-61; McKenzie, F., Ellison-Loschmann, L., Jeffreys, M. (2010) 'Investigating Reasons for Socioeconomic Inequalities in Breast Cancer Survival in New Zealand', *Cancer Epidemiology*, 34(6): 702-8; Pampalon, R., Hamel, D., Gamache, P., Raymond, G. (2009) 'A deprivation index for health planning in Canada', *Chronic Diseases in Canada*, 29(4): 178-191; Subramanian, S. V., Chen, J. T., Rehkopf, D. H., Waterman, P. D., Krieger, N. (2006) 'Comparing individual- and area-based socioeconomic measures for the surveillance of health disparities: A multilevel analysis of Massachusetts births, 1989-1991', *American Journal of Epidemiology*, 164(9): 823-34.

⁴⁸ Chowdry, H., Crawford, C., Dearden, L., Goodman, A., Vignoles, A. (2013) 'Widening Participation in Higher Education: Analysis Using Linked Administrative Data', *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 176(2): 431-457.

⁴⁹ See, for example, the CSO analysis of data from the Irish Health Survey at the website <http://www.cso.ie/en/releasesandpublications/ep/p-ihs/irishhealthsurvey2015/ct/> [accessed 12 October 2017].

⁵⁰ Osborne, B. and Shuttleworth, I. (2004) 'Widening Access to Higher Education in the UK: Questioning the Geographic Approach', *Higher Education Management and Policy*, 16(1) [DOI:10.1787/hemp-v16-art9-en].

⁵¹ Of course, it would be possible to argue that affluent individuals living in deprived areas are less strongly influenced by their social environment than deprived individuals living in affluent areas (a variant of the so-called 'ecological fallacy', which can arise when inferences are drawn about individuals using information on areas). This kind of difference is unlikely to be large enough to influence the results in a significant way but should be assessed during the implementation phase.

Another issue that emerges when using aggregate-level proxies relates to the stability of scores over time. One of the characteristics of the HP Deprivation Index is that it is based on statistical techniques which ensure that scores are comparable across successive waves of the census⁵². Of course, the characteristics of neighbourhoods may evolve over time due to changes in their composition (e.g. due to residential mobility) and/or changes in the situation of local residents (e.g. due to social mobility). In both cases, the (multivariate and multidimensional) structure of the HP Deprivation Index ensures that short-term fluctuations in the characteristics of neighbourhoods (e.g. changes in the female unemployment rate, for example) do not have an undue impact on their deprivation scores, whilst more significant processes of social mobility or gentrification are captured and tracked on a continuous scale.

We showed earlier that existing measures of socio-economic group based on occupational classifications can yield misleading estimates of inequalities in access to higher education due to the way in which occupations are sex-typed, due to changing family structures and due to the effects of social mobility. As the HP Deprivation Index captures (above all) the relativities between residential areas in terms of their population attributes, it is well-placed to overcome these difficulties. Rather than measuring the relative size of a set of categories, which are only roughly ordered from highest to lowest, we can analyse a continuous distribution of scores⁵³. It would be expected to correlate strongly with indicators of educational disadvantage, such as ‘first in family to attend higher education’ and other policy-relevant variables.

The indicators on which the HP Deprivation Index is based include membership of high and low social classes, but also educational attainments, demographic dynamics and labour market situation. By exploiting the strengths of composite indicators, this approach avoids a number of the shortcomings of alternative approaches. As they are measured on a continuous scale, deprivation scores can be divided into quintiles or deciles, and this provides a useful way of making comparisons over time and across different jurisdictions. They have a clear and unequivocal scale, and can be employed with great ease in funding formulae, eligibility criteria, weighting scales and other policy instruments.

Of course, the use of a Small Area deprivation index as a proxy for socio-economic position is not a panacea and cannot be expected to resolve all issues in relation to the assessment of inequalities in participation. It will therefore be important to assess its strengths and weaknesses ‘on the ground’ during the process of implementation.

⁵² Haase, T. and Pratschke, J. (2005) *Deprivation and its spatial articulation in the Republic of Ireland: new measures of deprivation based on the census of population 1991, 1996 and 2002*. Dublin: Pobal; Pratschke, J. and Haase, T. (2007) ‘Measurement of social disadvantage and its spatial articulation in the Republic of Ireland’, *Regional Studies*, 41(6): 719-734.

⁵³ See Bernard, J. (2006) ‘Promoting Access to Higher Education and Identifying Access Students: How Useful is Research on Participation by Socio-Economic Group?’ *International Studies in Sociology of Education*, 16(1), 19-35.

9. Measuring other attributes: students with a disability, Irish Travellers

Another challenge posed by the data plan involves showing how other categories – in addition to those defined by socio-economic position – can be identified using administrative data. The remaining target groups are (1) students with disabilities, (2) first-time, mature student entrants; (3) part-time/flexible learners; (4) further education award holders; (5) Irish Travellers. We will deal with categories (2), (3) and (4) in the next section, as they refer to categories of student, rather than identifiable subsets of the population. We will concentrate here on showing how students with disabilities and Irish Travellers might be identified within the new data plan.

Fortunately, the main educational databases already cover these two categories. For example, the Post-Primary Online Database contains a recently-introduced and voluntary measure of ethnic group, which currently covers roughly 70 per cent of students, and this includes a category for Irish Travellers. Although the quality of these fields needs to be assessed, and their coverage improved, linking with this data source would undoubtedly improve the monitoring of participation in higher education by members of this group.

As far as disabilities are concerned, it would be possible to count the number of students in receipt of support under the Fund for Students with Disabilities (FSD), perhaps in conjunction with or alongside information from the DARE scheme (20 HEIs participate in DARE, including all of the larger universities)⁵⁴. The HEA intends to integrate FSD data into its Student Record System from 2019, and this may provide an opportunity for obtaining an appropriate individual-level measure of disability.

Finally, the influence of family structure – another relevant target group that has been mentioned in several reports – on access to higher education could be estimated using aggregate data, although further work would be needed to evaluate this possibility and to assess the accuracy with which participation rates can be identified on the basis of spatial variations in the percentage of lone parent families, for example.

⁵⁴ Students are eligible for DARE if they have been diagnosed with one or more of the following: (1) Attention Deficit Disorder (ADD)/Attention Deficit Hyperactivity Disorder (ADHD); Autistic Spectrum Disorder (including Asperger's Syndrome); Blind/Vision Impaired; Deaf/Hearing Impaired; Developmental Coordination Disorder (DCD) – Dyspraxia/Dysgraphia; Mental Health Condition; Neurological Condition (including Brain Injury and Epilepsy); Physical Disability; Significant Ongoing Illness; Speech and Language Communication Disorder; Specific Learning Difficulty (including Dyslexia and Dyscalculia).

10. Other target groups: mature students, part-time students, further education award holders

It is important to identify mature students and to take their specific situation into account when studying equity of access. From a life course perspective, equity of access can be pursued by helping individuals from deprived backgrounds to make the transition from school to college, either directly or via deferred/delayed entry. Whilst deferred entry is likely to be a less effective strategy in labour market terms, it is nevertheless an important form of access. Indeed, deferred entry may represent a feasible way of reconciling the needs of people from deprived backgrounds and their aspirations regarding education and careers⁵⁵. From this perspective, deferred entry can gradually reduce the inequalities that are initially observed within a given cohort of school-leavers. In this sense, future access can at least partially 'redeem' the injuries inflicted by exclusion at an earlier age. For this reason, mature students may be seen as attempting to overcome educational disadvantage and arguably deserve support to do so.

Clancy (2015) notes that the Irish higher education system was traditionally characterised by a very low rate of participation among older students. This led policy-makers to assign a greater priority to encouraging mature students, following the publication of official reviews and reports in the 1990s. Comparative data now suggest that Ireland's situation is less anomalous than it used to be, with 13 per cent of new full-time undergraduate entrants being aged 23 and over in 2012.

From a measurement perspective, the integration of mature students within the data plan poses a challenge, as participation rates must be calculated in a specific way for this group. For example, if we analyse a group of college entrants, we will undoubtedly discover that they are of different ages and left school at different moments in the past. It would be misleading to add them all together and to express them as a proportion of just one cohort of school-leavers. It is more correct to confine the analysis to 18-20 year-olds (as the HEA currently does), although this only provides one part of the picture in relation to equity of access.

Alongside people who apply to enter higher education directly from school, there will always be some students who take a 'gap year', some who work for 2-3 years before enrolling, students who are returning to education after having children and students who have decided to change careers or lifestyle later on in life. The proportion of mature students can be quite significant, and their participation in higher education can have quite an important impact on macro-level inequalities. The implications of participation by mature students are distinct from those that derive from the participation of students making a direct transition.

⁵⁵ It is clearly important to distinguish between first-time mature students and those returning to higher education having previously interrupted their studies, although targeting both may be important from an access perspective. At the same time, not all mature students are from deprived backgrounds, which is why we suggested earlier that it is useful to distinguish between target groups in the population and forms of participation in higher education which may facilitate them.

Equity of access should arguably be assessed by age cohort. The key issue is whether the combination of policies adopted by various agencies and institutions can provide greater opportunities for participating in higher education, leading to a decline in inequalities. If an increasing number of mature students enter higher education, this may be seen as compensating for earlier inequalities (those observed when the respective age cohort left school and decided whether to apply to higher education, mitigated by subsequent re-entry to the education system). But it also contributes towards shaping a more equal education system and society.

Deferred entry to higher education as a mature student should therefore be modelled as a distribution across age groups. The resulting propensity to return to higher education will evolve under the influence of various factors, such as the economic cycle, cultural trends and government policy. We should therefore expect age, cohort and period effects to assert themselves over time. A rapid increase in participation rates within a specific social group may encourage older members of the same group to apply to higher education as mature students. This is similar to the net entry rate measure used by the OECD⁵⁶. Projected future rates of entry to higher education could be associated with current cohorts of school-leavers, allowing us to obtain a global assessment of equity of access which respects the differences across cohorts.

Turning now to part-time students and (more generally) those on ‘flexible learning’ programmes, the identification of these groups is straightforward and is already possible using the HEA’s Student Records System. As in the case of mature students, these are not social groups that can be identified within the population, which means that the individuals concerned cannot be targeted before they actually enrol. However, part-time enrolment may represent an attractive pathway for members of disadvantaged social groups, as it provides another way of reconciling participation in higher education with either paid employment or caring responsibilities, which is clearly of great importance from an equity of access perspective. Rather than describing mature and part-time students as ‘target groups’, it may be more appropriate to consider them as policy tools for encouraging participation by target groups. By facilitating enrolment as a flexible or mature student, and ensuring that adequate supports are in place to encourage these students, higher education institutions may be able to meet targets for widening participation more rapidly. This recognises that mature and part-time students do not always come from disadvantaged social groups, which may be important to bear in mind.

The last target group is students who are making the transition from further education. As in the case of mature students, part-time students and flexible learning programmes, this is considered a potentially important pathway for enabling students from disadvantaged and under-represented social groups to enrol in higher education. In the region of 6.5 per cent of

⁵⁶ Clancy (2015) provides a useful summary of how gross and net entry rates are currently measured, noting that the net entry rate for a specific age is obtained by dividing the number of first-time (new) entrants of that age by the total population of a corresponding age, multiplied by 100 (OECD, 2004a: 143). The overall net entry rate is calculated by summing the rates for each single year of age, providing an estimate of “the probability that a young person will enter tertiary education in his/her lifetime if current age-specific entry rates continue”.

students entering higher education each year do so on the basis of a further education qualification (see Table A7 in the National Plan 2015-19). Once again, people who have participated in further education do not constitute a target group in the strict sense of the term, and this category is likely to be quite heterogeneous. Nevertheless, it makes sense to monitor the number of people making the transition from further to higher education. This may involve an alternative access route or a straightforward application via the CAO⁵⁷ and it may enable people who initially chose a shorter and more vocationally-oriented course of study to progress to higher education.

⁵⁷ “In a report prepared for the Conference on Higher Education Selection and Entry Mechanisms, the Irish University Association (IUA, 2011) provides a summary of the evolution and current scale of the supplementary routes to accessing higher education that lie outside the main more competitive Central Applications Office (CAO) process. These schemes include entry routes for holders of FETAC awards, mature entry, the Higher Education Access Route (HEAR) for underrepresented socio-economic groups, and the Disability Access Route to Education (DARE). These access routes were taken by 5%, 16%, 2.5% and 0.9%, respectively, of all those who accepted a place in higher education in Ireland in 2010.” (Clancy, 2015, *op. cit.*)

11. Dealing with international flows

Another issue that should be addressed relates to the presence of foreign students at Irish HEIs and the tendency for a minority of Irish students to apply to UK universities. In fact, the number of students entering an Irish HEI in any given year does not provide a straightforward estimate of the probability of participating in higher education for the relevant age cohort, and this is not taken into account at the moment. A large inflow of foreign students would clearly inflate the estimated probability, whilst a large outflow of students attending foreign universities would reduce it. We should, ideally, control for both of these factors by excluding foreign students attending Irish HEIs from the numerator (as they cannot, by definition, be in the denominator) and by including Irish students attending foreign HEIs (as they are, by definition, included in the denominator).

In 2012/13, 1,340 Irish students were studying in a higher education institution in Northern Ireland and 6,955 were in Great Britain, which means that roughly 4 per cent of Irish students choose to study in the UK. It would be helpful, therefore, to include these students (many of whom would presumably be from relatively affluent families) by drawing on UCAS data to identify Irish students who accept a place in higher education in Great Britain or Northern Ireland⁵⁸.

The best way of doing this would be by using data from UCAS (which handles applications to UK higher education institutions) to identify Irish students who have accepted a place to study in Great Britain or Northern Ireland. The Administrative Data Research Centre in the UK has informed us that it is not currently possible to identify Irish students using data from UCAS, as they are included in a more general category (students from other EU countries). Alternative ways of obtaining data on Irish students in the UK should therefore be explored, focusing particularly on higher education institutions in Northern Ireland.

⁵⁸ The number of Irish students attending HEIs in other European and non-European countries is insignificant, by comparison.

12. Coverage of key datasets

We have identified eight sources of data that are particularly relevant to equity of access and could potentially contribute to the new data plan. These are (i) the Student Record System (**SRS**) developed with the support of the HEA, (ii) the Post-Primary Online Database (**P-POD**), (iii) the State Examination Commission (**SEC**), (iv) the **Census** of Population, (v) the Central Admissions Office (**CAO**) database on applications to higher education institutions, (vi) the Student Universal Support Ireland (**SUSI**) database on student contributions and maintenance grants, (vii) the Programme Learner Support System (**PLSS**) held by SOLAS and (viii) the UK Universities and College Admissions Services (**UCAS**) database on applications to higher education institutions (Great Britain and Northern Ireland). We will describe each of these datasets in greater detail below and indicate how they can contribute to the data plan.⁵⁹

(i) The HEA Student Record System (SRS) for students attending publicly-funded Higher Education Institutions⁶⁰

This system contains an individual record for each student in each of the HEA-funded institutions. The HEA has gathered data from the university and colleges sector since the 2004/2005 academic year, and from the Institutes of Technology since 2007/08. The SRS contains socio-economic data collected by HEIs as part of the *Equal Access Survey* and provides information on progression, transfers and completion of higher education qualifications.⁶¹

The HEA collects PPSNs, but these are not validated, and do not cover the entire student population. Up to 10-20% of students fail to provide valid PPSNs in some higher education institutions. Although the SRS does not currently contain the names and addresses of higher education students, the HEA has requested this information from HEIs, and is assessing ways of distinguishing between 'home' and 'term-time' addresses. The HEA then intends to utilise client identity services at the Department of Social Protection to verify and extend the coverage of PPSNs, once it has the appropriate information⁶². Internal matching is currently

⁵⁹ There are other sources of data on higher education students in Ireland, including the Irish Survey of Student Engagement (ISSE), the Eurostudent Survey and the School Leavers' Survey. These will not be discussed here, as the ISSE is not a sample survey (which makes it difficult to make reliable inferences), the Eurostudent Survey has a very low response rate (a little over 5% in 2013) and the School Leavers' Survey was discontinued in 2009. All three are very interesting data collection initiatives and merit greater attention than have so far received, although they are arguably less relevant to the current project than the other data sources described in this section.

⁶⁰ Victor Pigott from the HEA kindly provided information on the SRS (personal communication, June 2017).

⁶¹ HEA (2016) *A Study of Progression in Irish Higher Education 2012/13 to 2013/14*. Dublin: Higher Education Authority.

⁶² Files containing names, addresses, dates of birth and PPSNs can be sent to the Department of Social Protection for verification.

based on an internal student ID, which is different from the CAO application number, the Leaving Certificate exam number and the P-POD Pupil ID number.

(ii) The DES Post-Primary Online Database (P-POD) of school students

The Department of Education and Skills developed a (new) central database in 2014, containing up-to-date information on students and schools. Secondary schools are required to provide relevant information on students at the beginning of each year, and this information is used to allocate teaching posts and funding. The system includes Pupil identifiers and PPSN numbers,⁶³ residential addresses, date of birth, country of birth, ethnicity, nationality, progress through post-primary education and registration for state examinations.⁶⁴ Data on ethnicity/nationality do not currently have comprehensive coverage and their quality needs to be assessed. The current database has been geocoded in order to facilitate designations under the Delivering Opportunity in Schools (DEIS) programme, and this procedure will be maintained in the future.⁶⁵

(iii) The State Examination Commission (SEC) database of post-primary examination results⁶⁶

The SEC holds data on examination results for the Junior Certificate, Leaving Certificate and the Leaving Certificate Applied in line with Section 53 of the Education Act (1998). Examination results are (generally) linked to candidates by an examination number.⁶⁷ Existing legislation effectively excludes the possibility of assessing the comparative performance of schools in relation to academic achievement and means that any data that could be used to this end are not likely to be made available for analysis. However, the possibility of adding a field to P-POD, containing a coarsely-coded indicator of exam results at the end of post-primary education, should be explored. Alternatively, Leaving Certificate points are available within the SRS for candidates who applied through the CAO.

(iv) The Irish Census of Population

Held as a rule every five years, the Census of Population provides a wide range of aggregate-level demographic and socio-economic indicators. With the introduction of the new Small Area geography in the context of the 2011 census, the census has become even more

⁶³ See: <http://www.welfare.ie/en/Pages/Personal-Public-Service-Number-PPS-Number-Frequently-Asked.aspx>

⁶⁴ See: <http://www.education.ie/en/Schools-Colleges/Services/Returns/Post-Primary-Online-Database-P-POD-Project/>

⁶⁵ The consultants were directly involved in this process.

⁶⁶ Diarmuid Reidy from the DES kindly provided information on data held by the DES and other statutory bodies (personal communication, 12 June 2017).

⁶⁷ See, for example: <http://www.education.ie/en/Publications/Statistics/Education-Statistics-Database/>

important as a source of proxy data. Small Areas (SAs) comprise between 50 and 200 households and generally encompass either a complete neighbourhood or townland, or a relatively homogeneous part of one of these. Their relatively uniform populations and internal homogeneity⁶⁸ makes the SAs particularly useful as a source of proxy data.

SA-level deprivation scores are increasingly used to provide a proxy for individual socio-economic position. The existence of a consolidated and well-respected index of affluence and deprivation in Ireland – developed by the consultants – has encouraged this development. The Pobal HP Deprivation Index has the virtue of yielding scores which (i) cover the full spectrum from extreme affluence to extreme deprivation, (ii) follow an approximately normal distribution, (iii) are comparable across time, (iv) cover the distinctive urban and rural manifestations of social disadvantage and (v) are widely used and understood in administrative, research and policy-making circles. Several Government Departments and Agencies have adopted standard procedures involving geocoding beneficiaries' addresses and then linking them to Small Area deprivation scores, with a view to improving the targeting of their services, to develop resource allocation mechanisms or to evaluate outcomes.

(v) The Central Admissions Office (CAO) database of applications to Higher Education⁶⁹

In nearly all cases, applicants to the first year of undergraduate courses in higher education institutions in Ireland fill out an online CAO form, providing their name, address, Eircode, sex, date of birth, country of birth, nationality, school name/address, Leaving Certificate exam year and number (if relevant), other qualifications and choice of course. Applicants can apply for the HEAR and DARE schemes at the same time (and indicate whether they have applied for a maintenance grant or student contribution). This means that a considerable amount of information on these schemes (and others) is recorded on CAO servers. In addition, and where relevant, the system calculates the Leaving Certificate points for each course that an applicant applies to and any offers they are awarded and accept.

The CAO uses an individual application number to match individual data within its own databases. When it is necessary to match an application with external examination data, personal data are used, together with exam numbers and other numerical identifiers. The CAO receives the PPSN from the State Examination Commission (as part of the Leaving Certificate results file) although this is not normally used for matching. In addition, it collects PPSNs directly from applicants with a FETAC qualification (as QQI uses PPSNs) and from applicants to the HEAR scheme (as the DES and the HEA use PPSNs). The CAO recently began collecting Eircodes for applicants, although these do not have universal coverage and are not validated.

⁶⁸ This homogeneity is due to their small populations, as well as the algorithms used to define their boundaries.

⁶⁹ Kevin Keady from the CAO kindly provided information on their data systems (personal communication, June 2017).

(vi) The Student Universal Support Ireland (SUSI) database on student contributions and maintenance grants⁷⁰

The City of Dublin Education and Training Board (CDETb) holds data for higher and further education applicants who apply for a student contribution and/or maintenance grant by filling out an application form at the website www.susi.ie. Data collected include PPSN, address, Eircode (which is present for around 90% of new awards), age, gender, country of birth, nationality, residency, highest qualification level, course and Institution, mature student, own income (and income of partners, other relevant parties) as well as information on types of grant awarded.

SUSI has its own ID code for internal use (known as the ‘W number’), which is also used by colleges and students for administrative purposes. It is possible to distinguish between students who have been accepted for a course in another country (under 200 to Northern Ireland, less than 600 to the UK, and roughly 100 new entrants to other countries in 2016-17) and those who have been accepted by an Irish University, PLC, Institute of Technology or other institution. Although some information on One Parent Family payments is collected, this does not reliably identify grant recipients who are lone parents, or lone parents more generally. Mature students who apply for grants can be identified.

There is one precedent for linking individual-level data between SUSI and the SRS⁷¹. The identifying personal information present in the SUSI datasets is of high quality, as it is also used for carrying out cross-checks with the Tax Office or the Welfare Office. In terms of routine data sharing, this is essentially limited to gaining information that is necessary in order to process applications. There are seven categories of grant, which coincide with seven income thresholds, ranging from €23,000 (special maintenance grant) to €64,700 (50% student contribution). This is a precious source of data on family income which could complement other sources or be used to validate proxy indicators of socio-economic position.

The Programme Learner Support System (PLSS) containing data on students in Further Education and Training held by SOLAS⁷²

The PLSS was instituted in the second half of 2013, building on work by the SOLAS Implementation Group in 2012. This important development was motivated by the need for

⁷⁰ Vincent Downey at CDETb kindly provided information on datasets maintained by this organisation (personal communication, July 2017).

⁷¹ HEA (2015b) *Student Grant Recipients from a First Year Full-Time Undergraduate New Entrant Cohort for the Academic Year 2013/14 in HEA Funded Institutions*. Dublin: Higher Education Authority. This study shows that across all HEA-funded institutions, 42% of new entrants are in receipt of full fees, with various rates of maintenance grant.

⁷² Tom Craig from SOLAS kindly provided information on the PLSS system (personal communication, 26 October 2017) See also:
<https://www.qqi.ie/Downloads/Programme%20Learner%20Support%20System%20EQU%20VET%20-%20Fiona%20Maloney%20ETBI.pdf>

systematic and timely access to data on Further Education and Training provision, including enrolment, drop-out, completion and certification. The system contains information on courses, learning programmes and awards and the Learner Database contains information on student characteristics. The PLSS is now fully developed and operational, and uses the PPSN for individual identification. If data protection criteria can be satisfied, this would be an appropriate source of information on Further Education and Training, with a view to assessing educational inequalities more generally and for studying progression from further to higher education.

(vii) The UK Universities and College Admissions Services (UCAS) database on students attending Higher Education Institutions

This database contains individual-level information on applications to higher education institutions in Great Britain and Northern Ireland (equivalent to the CAO). A considerable number of Irish school students apply to such HEIs each year, and it is worthwhile pursuing the possibility of identifying them and including them in the analysis of access. UCAS is accustomed to providing data for research and other purposes⁷³, including information on HEI, type of course, country of domicile, acceptance route, age, gender, ethnic group and examination results and routinely reports the breakdown of applicants by the deprivation score of their small area of residence⁷⁴. The prospect of identifying students who go to UK HEIs is an attractive one, as they are likely to be from relatively affluent backgrounds, and the failure to identify them as HE entrants could give rise to bias. UCAS recently made nine years of individual-level university and college admissions data accessible to researchers through the Administrative Data Research Network. This decision was motivated by the need to improve understanding of equality of access to higher education. However, cross-border data-linking exercises are not currently possible via the ADRN, and the available UCAS data cannot be used to identify Irish students.

⁷³ See: <https://wwwucas.com/corporate/data-and-analysis>.

⁷⁴ UCAS Analysis and Research (2015) 50 per cent increase in demand for higher education from the most deprived areas of Scotland since 2006. UCAS Analysis note 2015/03.

13. Developing a Shared Research Agenda

An essential but often-overlooked component of a good data plan is a discussion of how the data will be used in order to produce an evidence base for policy-makers. Data never “speak for themselves”, but are analysed through the prism of existing theories, measures and models. Certain statistical models have quite rigid requirements in relation to data – for example, in relation to the units of analysis and the definition of key indicators or covariates – and it is useful to build an awareness of these requirements into the design of the new data plan. A failure to anticipate the kind of data structures that might be needed in order to monitor and assess equity of access to educational opportunities in the future could lead to criticism and disillusionment.

As we noted earlier, the research questions that are being posed across different levels of the education system are increasingly inseparable, and demand a systemic response. For example, the evaluation of the DEIS programme would inevitably touch on higher education enrolments, whilst an evaluation of the HEAR initiative would inevitably require data on participation in post-primary education and attainments at the Leaving Certificate. Policy makers are increasingly aware of the need to reassemble these different pieces of the picture and to develop a research agenda which has the necessary systemic and holistic focus. A useful starting-point when seeking to assess the impact of policies is to build a combined model of educational inequalities, and then to imagine interventions as seeking to increase the probabilities of transition for disadvantaged groups. An example of this kind of theoretical model is shown in Figure 2 below.

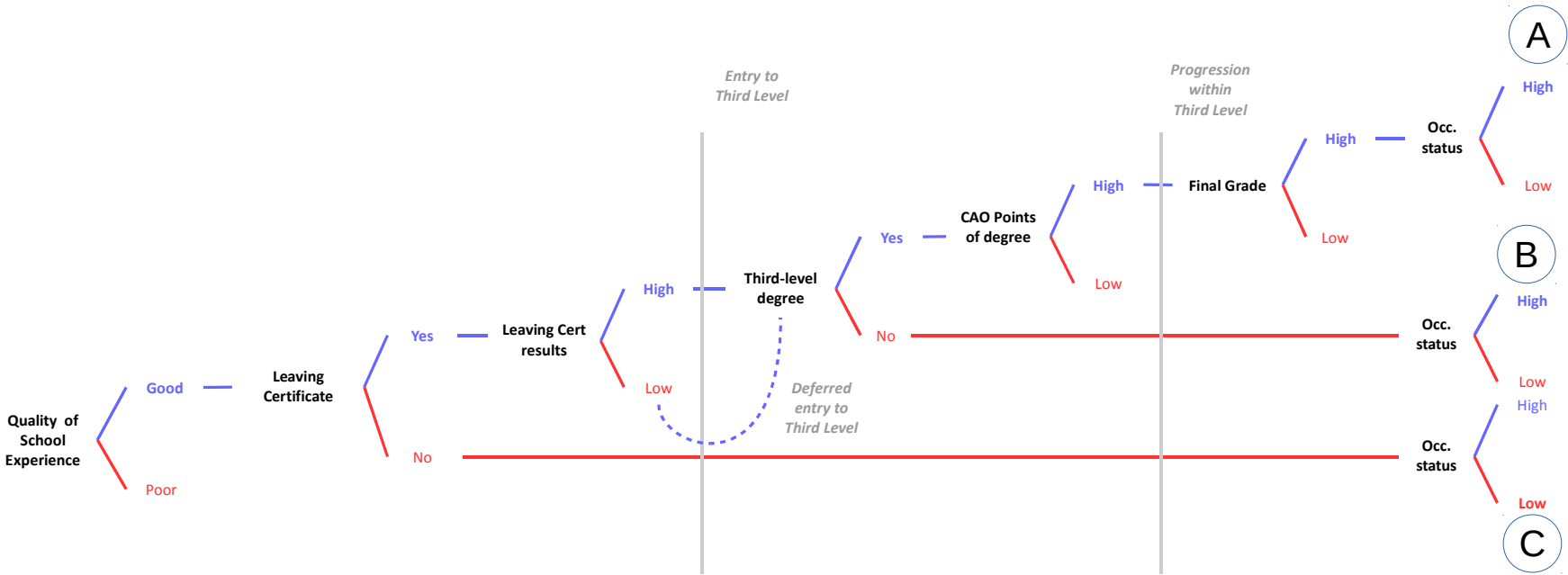
The figure illustrates three different kinds of educational trajectories (A, B and C), and shows direct and indirect paths from school to university. 'A' represents an 'élite' trajectory, culminating in the achievement of a high-status professional position. 'B' reflects the achievement of high occupational status without attending university, whilst 'C' reflects the experience of a student who fails to complete the Leaving Certificate and has no alternative but to accept a low-skilled job.

The probability of completing each transition shown in the figure is influenced not only by the characteristics of the individual, their family and the neighbourhood, but also by their previous transitions and experiences, giving rise to an incremental process. For example, access to higher education is dependent on applying to a HEI and on achieving the required points, which is dependent on Leaving Cert results, which are dependent on school experience and so on. This means that disadvantage accumulates over time and inequalities can be distributed unevenly within the education system.

Inequalities in relation to one transition may be explained by previous transitions, as when the under-representation of a particular social group at university is due to a tendency to leave school before completing the Leaving Certificate. It is also important to stress the difference between inequalities that are due to explicit selection mechanisms and inequalities that are due to choices or informal selections. For example, the under-representation of people from disadvantaged areas, within the university system, may be due (at least in part) to the difficulties they face in meeting entrance requirements or they could be due to a tendency to seek paid employment after completing school.

Figure 2

Model of educational transitions



Analyses of educational inequalities may be either descriptive or explanatory in their aims. In the first case, tables and graphs are used to compare transition rates across categories, taken either singly or in combination. The tables presented in the report by O'Connell *et al.* (2006) are a good example of a descriptive analysis of socio-economic disparities in participation.

In the second case, statistical models are used in an attempt to explain differences in the propensity to participate in education and to pursue specific kinds of educational attainments. For example, logistic regression models might be used to identify characteristics that are associated with the decision to enrol in higher education or event history models might be used to study the timing of this transition. The statistical analyses presented in the report by McCoy *et al.* (2010), based on EU-SILC data, are a good example of this type of quantitative analysis. More sophisticated models, such as those which take account of the nesting of pupils within schools and neighbourhoods, typically require additional data on these higher-level units, as well as the possibility of associating individuals with schools, neighbourhoods and other contexts.

By enabling researchers to work with individual-level data, and by providing access to all necessary continuous and categorical indicators, all of these analyses can be carried out under the new data plan. Where additional information are required, linking at the individual level, at the aggregate level, or by Small Area remains the most powerful way of integrating the necessary data. By creating geocoded datasets, it will also be possible to identify geographical areas that have low rates of participation in higher education, even after controlling for socio-economic and demographic composition, which is another objective of the National Plan⁷⁵. The most important consideration, therefore, in relation to statistical analysis, is the need for full access to linked data within a secure environment.

It would, of course, be self-defeating to construct a powerful linked dataset dealing with educational inequalities and then refuse to allow researchers to exploit its full potential. Whether they have been commissioned by one of the organisations providing data, by central government or whether they are working on their own research projects, *bona fide* research using the linked datasets should be permitted. This means that appropriate criteria must be developed for awarding access, adequate protocols should be put in place to ensure data security and anonymity during this process, and the results should be made public. It would, moreover, be appropriate to include researchers in discussions of the new data plan, as they represent another important stakeholder. When identifying a 'trusted third party' for the purposes of data linking, therefore, some consideration should be given to how research access is to be provided.

⁷⁵ HEA (2015b) *op. cit.*, p. 37.

14. Change Management

In this section, we describe the main steps in the implementation of the new framework, indicating the agencies involved, the amount of time required and any other inputs or resources that may be necessary (e.g. data, identifiers, funds, specialist services, skills). Each element in the final plan will form part of a coherent implementation strategy.

A crucial component of the change management plan is the way in which it distinguishes between long-term strategic objectives (e.g. involving ambitious inter-agency data linkage) and short-term requirements (e.g. improving measurements and estimates to facilitate analysis of equity in access to higher education). In this proposal, we have shown that these objectives are in harmony, and that the HEA can work simultaneously towards short-term improvements in measures of access and medium-term improvements in the scope of the available data. The change management strategy is divided into the following phases:

Phase 1 Preparatory work

During the first phase, the HEA will promote dialogue between stakeholders with a view to building a consensus around the core elements of the data plan: sharing information and collaborating on the development of a new 'data infrastructure' for the education sector as a whole. As part of this process, the HEA will, alongside and together with other organisations, seek to identify key enablers, including the pursuit of appropriate primary legislation that can motivate and legitimate the sharing of data between Departments, State Agencies and other organisations involved in the delivery of education, in order to develop, monitor or assess policies to promote equality in access to education.

The following organisations should be invited to participate in this process, together with the HEA: DES, CAO, SEC, CDETB, SOLAS, CSO. A key aim of Phase 1 is to achieve a collective understanding of the benefits and advantages offered by the data plan to all of the participating organisations, as well as to the population more generally. In this phase, the data plan must evolve from being a HEA proposal to being a joint undertaking by all of the aforementioned stakeholders, implying that organisational roles and responsibilities must be discussed and agreed collectively. This phase is expected to require at least 6 months.

Phase 2 Initial convergence in procedures

During this phase, stakeholders will return to their own systems and implement harmonised procedures for the identification of individuals, for the collection of any additional data that may be required and in order to improve the quality and coverage of key variables. This is a necessary initial step in order to share information, and must be applied on the basis of shared definitions and protocols. All information to be collected must be defined in a particularly comprehensive and precise way, so that names, 'home' and 'term-time' addresses, dates of birth and nationalities are appropriately collected and validated with PPSNs and Eircodes. As noted earlier, considerable progress has already been made in relation to this aim within both the DES and the HEA.

It is important that there be a degree of redundancy in the collection of identifying personal information, as this will facilitate additional checks. Procedures for identifying mistakes and correcting them should also be agreed. At the end of this phase, all participating organisations will have revised their datasets and data collection procedures in order to ensure that individuals can be identified in a predictable and unequivocal manner. Where key pieces of information for the assessment of equity of access are found to be missing, agreement must also be reached on how to fill these gaps. This will lead to the preparation of a joint codebook specifying the variables and formats used in each dataset.

A data protection impact assessment will be commissioned during this phase of the project, with a view to satisfying legal requirements and providing firm foundations for the new procedures for measuring equity of access. This will enable the HEA to identify the requirements of the new plan in terms of enabling legislation. This phase will overlap with Phase 1 and can be completed within 18-24 months.

Phase 3 Development of agreements and protocols

The aim of the second phase is to set out the conditions, protocols and procedures for linking data and analysing merged data on educational inequalities. These protocols and procedures will be based on the new EU and national rules regarding data protection, as well as established rules and mechanisms for exchanging data within the public sector. The main agreements and protocols for joining datasets will be assessed by specialists in data protection in order to ensure that they are in line with the prevailing norms⁷⁶. During this phase, formal agreements will be signed between the HEA, DES and other participating organisations. An appropriate 'trusted third party' will be identified and terms and conditions for accessing data will be defined. This phase will partly overlap with the first two, and can be completed within two years.

Phase 4 Geocoding

The fourth phase will be concerned with geocoding the addresses of students in each dataset. This may have to be carried out autonomously by each organisation using its own resources. At the end of this phase, all organisations will be in a position to share aggregate count data based on commonly-agreed criteria and specifications. They will also be in a position to analyse their own data using Small Area deprivation scores. This phase will require between 3 and 6 months, and is dependent upon successful completion of Phase 2.

⁷⁶ A key issue in this context is the EU General Data Protection Regulation (GDPR), which has already been approved by the European Parliament and will be enforced from 25 May 2018. This regulation will be accompanied and complemented by national legislation, and together these norms will enable researchers to join and analyse individual-level data only under specific and binding conditions. A legislative mandate will typically be required for data collection by public bodies and the people involved must be informed (or otherwise understand) that the information they provide can be used for research purposes. Naturally, anonymity will also have to be assured. This means that public bodies with a 'legitimate basis' for collecting personal data at present will generally require an 'explicit legislative mandate' to do so.

Phase 5 Data linking

The fifth phase will involve linking data on individuals across datasets using common identifiers. To ensure respect for data protection regulations, each case should be associated with a pseudo-anonymous identifier before leaving the host organisation, or the dataset could be encrypted before being merged with the other datasets by a trusted third party organisation. This organisation could be either public or private, and would have the task of linking data from different sources, either within a protected data centre or subject to assessment of data security procedures at other locations.

As noted earlier, the procedures for linking data will be developed during Phases 1 and 2, and will be agreed formally during Phase 3. This means that data linking during this phase will not require more than 3 months. During Phase 5, efforts will also be made to extend data linking to new sources, including SOLAS, UCAS, CAO and SUSI. This will facilitate more accurate controls regarding routes to higher education as well as providing data that can be used to assess and complement the Small Area proxy measure of socio-economic position discussed earlier.

Phase 6 Analysis

During the sixth and final phase, the required analyses of equity in access will be carried out. Monitoring activities may be carried out each year, or at regular intervals, and involve targeted analysis of single issues in relation to equity of access or broader analysis of a set of themes, on behalf of organisations in the network or groups of organisations. Applicants should be required to submit a written proposal for approval by data holders, in which they set out the aims of the analysis, the techniques and methods to be used, and the kind of output and results to be published or circulated. It would be possible to imagine a role for the CSO as trusted third party⁷⁷. Any questions or difficulties encountered during the analysis could be communicated to the participants in the network, who could meet at regular intervals to discuss ways of maintaining data quality, extending their cooperation and providing policy-relevant information to interested parties.

⁷⁷ The CSO is currently playing this role in the context of another HEA project, in which data on higher education students are linked to earnings data from the Revenue Commissioner and information on employment situation from the Department of Social Protection, using PPS numbers.

15. Conclusions

In this report, we have described the essential contours of the new data plan for monitoring and assessing equity of access to higher education. This data plan should be seen as forming part of a joint analytical undertaking involving several organisations, in which *linked datasets are developed and used to satisfy the requirements of those bodies in relation to social inclusion monitoring in education*. This may be expected to yield considerable benefits in terms of economies of scale, efficiency and opportunities for cooperation as well as promoting an integrated approach to policy-making and policy delivery.

There are two strong arguments for the data plan, the first of which is based on the importance of promoting equality of opportunity from a “social justice” perspective, and the second of which is grounded in the greater effectiveness and efficiency associated with using merged administrative datasets. It will be necessary to convince all parties that data linking and sharing is not only useful and desirable, but absolutely essential in order to fulfil statutory obligations. The reduction of inequalities in access to higher education is a key policy objective and is fundamental to achieving greater social justice, improving the competitiveness of the national economy, contributing to the sustainability and effectiveness of the higher education system itself and maintaining social cohesion. When discussing the more technical aspects of the data plan, it is important not to lose sight of the enormous substantive importance of equality of access as a strategic objective. We suggested that the move towards increasingly integrated, joined-up and “smart” policies and services must be accompanied by similarly integrated, joined-up and inter-sectoral data initiatives which have the potential to provide knowledge to guide policy-making in the future⁷⁸.

The fact that this approach challenges established procedures, as well as requiring a sophisticated approach to data protection, does not mean that it is impractical or unachievable. Rather than steering away from this promising research frontier, the HEA and other organisations committed to widening educational participation must find ways of moving towards it. This can create powerful “win/win” synergies across different sectors where each step towards the optimal data situation generates new analytical opportunities and contributes to overcoming obstacles.

The first step in this process is arguably finding a way to share and link information from P-POD, on the one hand, and the SRS, on the other. This would yield considerable benefits for both the DES and the HEA and would contribute to a far-reaching reformulation of data requirements for future monitoring activities. Any obstacles to this should be addressed in an open and confident way, with a view to identifying practical solutions. Individual-level data on clients and citizens are routinely linked, merged, exchanged and analysed throughout the Irish public sector in order to fulfil the statutory role of organisations like the HEA, the CDETB and the DES. Indeed, it is difficult to imagine an area of policy delivery that would not be severely affected by interruptions in this kind of flow of individual data. As

⁷⁸ The National Plan argues that “Equity of access policies must span the entire education spectrum and take a ‘whole of education’ approach to social inclusion” (HEA, 2015a, *op. cit.*, p. 16).

policy-relevant research and assessment are typically not part of routine statutory obligations, research based on data linking have come to be seen as problematic in terms of data protection. Whilst seeking to overcome resistance to this plan, it arguably makes sense to pursue legislative change which can provide a solid normative foundation for this research agenda.

It is helpful to think of this new data plan as contributing to the national data infrastructure, we argued earlier, as this enables us to appreciate how the proposed data plan relates to other datasets and information systems. This infrastructure comprises a number of different elements, ranging from legislation to data protocols, public and private organisations, new and existing datasets and statistical definitions. The data plan is about bringing these different elements into a specific kind of configuration in order to provide policy-makers and administrators with reliable and accurate information on educational equality, and access to higher education in particular, with a view to improving participation for target groups mentioned in national plan. The plan includes a realistic assessment of how to achieve this goal, by mapping out the landscape in relation to educational data, identifying potential ‘roadblocks’, on the one hand, and ‘enabling factors’, on the other.

In this data plan, we have described how the national data infrastructure might be developed in order to achieve national strategic objectives, but it is important for real progress in analysing and understanding educational inequalities to be made at each step. Thus, the plan will yield benefits to Government and stakeholders immediately, not when it has been fully completed. It is an incremental plan which will adapt itself to the prevailing context and will evolve over time on the basis of the available opportunities for data sharing and linking. It comprises minimalist and maximalist definitions which provide useful points of reference for a process that will undoubtedly involve compromise and flexibility on the part of all stakeholders.

16. Bibliography

- Bernard, J. (2006) 'Promoting Access to Higher Education and Identifying Access Students: How Useful is Research on Participation by Socio-Economic Group?' *International Studies in Sociology of Education*, 16(1), 19-35.
- Bologna Process (2007) *Towards the European Higher Education Area: responding to challenges in a globalised world*. Communiqué of the Conference of European Ministers Responsible for Higher Education, 18 May 2007, London, available at http://media.ehea.info/file/2007_London/69/7/2007_London_Communique_English_588697.pdf, accessed November 2017.
- Byrne, D., Doris, A., Sweetman, O. and Casey, R. (2013) *An Evaluation of the HEAR and DARE Supplementary Admission Routes to Higher Education*. Maynooth: NUI Maynooth.
- Byrne, D. and Murray, C. (2017) *An Independent Review to Identify the Supports and Barriers for Lone Parents in Accessing Higher Education and to Examine Measures to Increase Participation*. Maynooth: Maynooth University.
- Chowdry, H., Crawford, C., Dearden, L., Goodman, A. and Vignoles, A. (2013) 'Widening Participation in Higher Education: Analysis Using Linked Administrative Data', *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 176(2): 431–457.
- Clancy, P. (1982) *Participation in Higher Education*. Dublin: Higher Education Authority.
- Clancy, P. (1988) *Who Goes to College?* Dublin: Higher Education Authority.
- Clancy, P. (1995) *Access to College: Patterns of Continuity and Change*. Dublin: Higher Education Authority.
- Clancy, P. (2001) *College Entry in Focus: A Fourth National Survey of Access to Higher Education*. Dublin: Higher Education Authority.
- Clancy, P. (2015) *Irish Higher Education: A Comparative Perspective*. Dublin: IPA.
- Clancy, P., Hayden, J., Kelleher, M. (2010) *External Audit of Equal Access Survey*. Dublin: Higher Education Authority.
- Domínguez-Berjón, F. (2005) 'The Usefulness of Area-Based Socioeconomic Measures to Monitor Social Inequalities in Health in Southern Europe', *European Journal of Public Health*, 16(1): 54-61.
- Erikson, R. (1984) 'Social Class of Men, Women and Families', *Sociology*, 18, 4: 500-514.
- Fleming, T., Loxley, A. and Finnegan, F. (2017) *Access and Participation in Irish Higher Education*. London: Palgrave Macmillan.
- Forsyth, A. and Furlong, A. (2000) *Socioeconomic Disadvantage and Access to Higher Education*. Bristol: Policy Press.
- Haase, T. and Pratschke, J. (2005) *Deprivation and its spatial articulation in the Republic of Ireland: new measures of deprivation based on the census of population 1991, 1996 and 2002*. Dublin: Pobal
- HEA (2010) *National Plan to Achieve Equity of Access to Higher Education 2008-2013: Mid-term Review*. Dublin: Higher Education Authority.

- HEA (2015a) *National Plan for Equity of Access to Higher Education 2015-2019*. Dublin: Higher Education Authority.
- HEA (2015b) *Student Grant Recipients from a First Year Full-Time Undergraduate New Entrant Cohort for the Academic Year 2013/14 in HEA Funded Institutions*. Dublin: Higher Education Authority.
- HEA (2016) *A Study of Progression in Irish Higher Education, 2012/13 to 2013/14*. Dublin: Higher Education Authority.
- HEA (2016) *Higher Education System Performance 2014–2016*. Dublin: Higher Education Authority.
- HEA (2017a) *Key Facts and Figures. Higher Education 2015-16*. Dublin: Higher Education Authority.
- HEA (2017b) *What Do Graduates Do? The Class of 2015*. Dublin: HEA.
- IUA (2011) *Higher Education Selection and Entry Mechanisms: Issues and Options, Report of the IUA Working Group*. Dublin: Irish Universities Association.
- Kooij, Y. (2015) *European Higher Education Policy and the Social Dimension: A Comparative Study of the Bologna Process*. London: Palgrave Macmillan.
- McCoy, S., Byrne, D., O’Connell, P., Kelly, E. and Doherty, C. (2010) *Hidden Disadvantage? A Study on the Low Participation in Higher Education by the Non-Manual Group*. Dublin: Higher Education Authority.
- McKenzie, F., Ellison-Loschmann, L. and Jeffreys, M. (2010) ‘Investigating Reasons for Socioeconomic Inequalities in Breast Cancer Survival in New Zealand’, *Cancer Epidemiology*, 34(6): 702-8
- O’Connell, P., Clancy, D. and McCoy, S. (2006) *Who Went to College in 2004? A National Survey of New Entrants to Higher Education*. Dublin: Higher Education Authority.
- OECD 2004. *OECD Handbook for Internationally Comparative Education Statistics*. Paris: OECD.
- OECD 2009. *OECD Reviews of Migrant Education*. Ireland. Paris: OECD.
- Osborne, B. and Shuttleworth, I. (2004) ‘Widening Access to Higher Education in the UK: Questioning the Geographic Approach’, *Higher Education Management and Policy*, 16(1) [DOI:10.1787/hemp-v16-art9-en].
- Pampalon, R., Hamel, D., Gamache, P. and Raymond, G. (2009) ‘A deprivation index for health planning in Canada’, *Chronic Diseases in Canada*, 29(4): 178-191
- Pratschke, J. and Haase, T. (2007) ‘Measurement of social disadvantage and its spatial articulation in the Republic of Ireland’, *Regional Studies*, 41(6): 719-734.
- Read, P., Higgs, G. and Taylor, G. (2005) ‘The Potential and Barriers to the Use of Geographical Information Systems for Marketing Applications in Higher Educational Institutions’. *Marketing Intelligence & Planning*, 23 (1): 30-42.
- Subramanian, S. V., Chen, J. T., Rehkopf, D. H., Waterman, P. D. and Krieger, N. (2006) ‘Comparing individual- and area-based socioeconomic measures for the surveillance of health disparities: A multilevel analysis of Massachusetts births, 1989-1991’, *American Journal of Epidemiology*, 164(9): 823-34.